Titre: Title:	Automatic Segmentation of Intramedullary Multiple Sclerosis Lesions
Auteur: Author:	Charley Gros
Date:	2018
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Gros, C. (2018). Automatic Segmentation of Intramedullary Multiple Sclerosis Lesions [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/3200/

UNIVERSITÉ DE MONTRÉAL

AUTOMATIC SEGMENTATION OF INTRAMEDULLARY MULTIPLE SCLEROSIS LESIONS

CHARLEY GROS

INSTITUT DE GÉNIE BIOMÉDICAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE BIOMÉDICAL)

MAI 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

AUTOMATIC SEGMENTATION OF INTRAMEDULLARY MULTIPLE SCLEROSIS LESIONS

présenté par : GROS Charley

en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées

a été dûment accepté par le jury d'examen constitué de :

M. BILODEAU Guillaume-Alexandre, Ph. D, président

M. COHEN-ADAD Julien, Ph. D, membre et directeur de recherche

M. BROWN Robert A., MD, Ph. D, membre

DEDICATION

À ma mère, Brigitte, et à mon père, Jean-Pierre.

RÉSUMÉ

Contexte: La moelle épinière est un composant essentiel du système nerveux central. Elle contient des neurones responsables d'importantes fonctionnalités et assure la transmission d'informations motrices et sensorielles entre le cerveau et le système nerveux périphérique. Un endommagement de la moelle épinière, causé par un choc ou une maladie neurodégénérative, peut mener à un sérieux handicap, pouvant entraîner des incapacités fonctionnelles, de la paralysie et/ou de la douleur. Chez les patients atteints de sclérose en plaques (SEP), la moelle épinière est fréquemment affectée par de l'atrophie et/ou des lésions. L'imagerie par résonance magnétique (IRM) conventionnelle est largement utilisée par des chercheurs et des cliniciens pour évaluer et caractériser, de façon non-invasive, des altérations micro-structurelles. Une évaluation quantitative des atteintes structurelles portées à la moelle épinière (e.g. sévérité de l'atrophie, extension des lésions) est essentielle pour le diagnostic, le pronostic et la supervision sur le long terme de maladies, telles que la SEP. De plus, le développement de biomarqueurs impartiaux est indispensable pour évaluer l'effet de nouveaux traitements thérapeutiques. La segmentation de la moelle épinière et des lésions intramédullaires de SEP sont, par conséquent, pertinentes d'un point de vue clinique, aussi bien qu'une étape nécessaire vers l'interprétation d'images RM multiparamétriques. Cependant, la segmentation manuelle est une tâche extrêmement chronophage, fastidieuse et sujette à des variations inter- et intra-expert. Il y a par conséquent un besoin d'automatiser les méthodes de segmentations, ce qui pourrait faciliter l'efficacité procédures d'analyses. La segmentation automatique de lésions est compliqué pour plusieurs raisons: (i) la variabilité des lésions en termes de forme, taille et position, (ii) les contours des lésions sont la plupart du temps difficilement discernables, (iii) l'intensité des lésions sur des images MR sont similaires à celles de structures visiblement saines. En plus de cela, réaliser une segmentation rigoureuse sur l'ensemble d'une base de données multi-centrique d'IRM est rendue difficile par l'importante variabilité des protocoles d'acquisition (e.g. résolution, orientation, champ de vue de l'image). Malgré de considérables récents développements dans le traitement d'images MR de moelle épinière, il n'y a toujours pas de méthode disponible pouvant fournir une segmentation rigoureuse et fiable de la moelle épinière pour un large spectre de pathologies et de protocoles d'acquisition. Concernant les lésions

intramédullaires, une recherche approfondie dans la littérature n'a pas pu fournir une méthode disponible de segmentation automatique.

Objectif: Développer un système complètement automatique pour segmenter la moelle épinière et les lésions intramédullaires sur des IRM conventionnelles humaines.

Méthode: L'approche présentée est basée de deux réseaux de neurones à convolution mis en cascade. La méthode a été pensée pour faire face aux principaux obstacles que présentent les données IRM de moelle épinière. Le procédé de segmentation a été entrainé et validé sur une base de données privée composée de 1943 images, acquises dans 30 différents centres avec des protocoles hétérogènes. Les sujets scannés comportent 459 sujets sains, 471 patients SEP et 112 avec d'autres pathologies affectant la moelle épinière. Le module de segmentation de la moelle épinière a été comparé à une méthode existante reconnue par la communauté, *PropSeg*.

Résultats: L'approche basée sur les réseaux de neurones à convolution a fourni de meilleurs résultats que *PropSeg*, atteignant un Dice médian (intervalle inter-quartiles) de 94.6 (4.6) vs. 87.9 (18.3) %. Pour les lésions, notre segmentation automatique a permis d'obtenir un Dice de 60.0 (21.4) % en le comparant à la segmentation manuelle, un ratio de vrai positifs de 83 (34) %, et une précision de 77 (44) %.

Conclusion: Une méthode complètement automatique et innovante pour segmenter la moelle épinière et les lésions SEP intramédullaires sur des données IRM a été conçue durant ce projet de maîtrise. La méthode a été abondamment validée sur une base de données clinique. La robustesse de la méthode de segmentation de moelle épinière a été démontrée, même sur des cas pathologiques. Concernant la segmentation des lésions, les résultats sont encourageants, malgré un taux de faux positifs relativement élevé. Je crois en l'impact que peut potentiellement avoir ces outils pour la communauté de chercheurs. Dans cette optique, les méthodes ont été intégrées et documentées dans un logiciel en accès-ouvert, la "Spinal Cord Toolbox". Certains des outils développés pendant ce projet de Maîtrise sont déjà utilisés par des analyses d'études cliniques, portant sur des patients SEP et sclérose latérale amyotrophique.

ABSTRACT

Context: The spinal cord is a key component of the central nervous system, which contains neurons responsible for complex functions, and ensures the conduction of motor and sensory information between the brain and the peripheral nervous system. Damage to the spinal cord, through trauma or neurodegenerative diseases, can lead to severe impairment, including functional disabilities, paralysis and/or pain. In multiple sclerosis (MS) patients, the spinal cord is frequently affected by atrophy and/or lesions. Conventional magnetic resonance imaging (MRI) is widely used by researchers and clinicians to non-invasively assess and characterize spinal cord microstructural changes. Quantitative assessment of the structural damage to the spinal cord (e.g. atrophy severity, lesion extent) is essential for the diagnosis, prognosis and longitudinal monitoring of diseases, such as MS. Furthermore, the development of objective biomarkers is essential to evaluate the effect of new therapeutic treatments. Spinal cord and intramedullary MS lesions segmentation is consequently clinically relevant, as well as a necessary step towards the interpretation of multi-parametric MR images. However, manual segmentation is highly time-consuming, tedious and prone to intra- and inter-rater variability. There is therefore a need for automated segmentation methods to facilitate the efficiency of analysis pipelines. Automatic lesion segmentation is challenging for various reasons: (i) lesion variability in terms of shape, size and location, (ii) lesion boundaries are most of the time not well defined, (iii) lesion intensities on MR data are confounding with those of normal-appearing structures. Moreover, achieving robust segmentation across multi-center MRI data is challenging because of the broad variability of data features (e.g. resolution, orientation, field of view). Despite recent substantial developments in spinal cord MRI processing, there is still no method available that can yield robust and reliable spinal cord segmentation across the very diverse spinal pathologies and data features. Regarding the intramedullary lesions, a thorough search of the relevant literature did not yield available method of automatic segmentation.

Goal: To develop a fully-automatic framework for segmenting the spinal cord and intramedullary MS lesions from conventional human MRI data.

Method: The presented approach is based on a cascade of two Convolutional Neural Networks (CNN). The method has been designed to face the main challenges of 'real world' spinal cord MRI data. It was trained and validated on a private dataset made up of 1943 MR volumes, acquired in different 30 sites with heterogeneous acquisition protocols. Scanned subjects involve 459 healthy controls, 471 MS patients and 112 with other spinal pathologies. The proposed spinal cord segmentation method was compared to a state-of-the-art spinal cord segmentation method, *PropSeg*.

Results: The CNN-based approach achieved better results than *PropSeg*, yielding a median (interquartile range) Dice of 94.6 (4.6) vs. 87.9 (18.3) % when compared to the manual segmentation. For the lesion segmentation task, our method provided a median Dice-overlap with the manual segmentation of 60.0 (21.4) %, a lesion-based true positive rate of 83 (34) % and a lesion-based precision de 77 (44) %.

Conclusion: An original fully-automatic method to segment the spinal cord and intramedullary MS lesions on MRI data has been devised during this Master's project. The method was validated extensively against a clinical dataset. The robustness of the spinal cord segmentation has been demonstrated, even on challenging pathological cases. Regarding the lesion segmentation, the results are encouraging despite the fairly high false positive rate. I believe in the potential value of these developed tools for the research community. In this vein, the methods are integrated and documented into an open-source software, the Spinal Cord Toolbox. Some of the tools developed during this Master's project are already integrated into automated analysis pipelines of clinical studies, including MS and Amyotrophic Lateral Sclerosis patients.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ALS	amyotrophic lateral sclerosis
ALSFRS	amyotrophic lateral sclerosis functional rating scale
CIS	clinically isolated syndrome
CNN	convolutional neural networks
CNS	central nervous system
CRF	conditional random fields
CSF	cerebrospinal fluid
DC	dorsal column
FOV	field of view
GLCM	gray level co-occurrence matrix
HoG	histogram of oriented gradients
LF	lateral funiculus
MRF:	Markov random fields
MRI	magnetic resonance imaging
MS	multiple sclerosis
PNS	peripheral nervous system
PVE	partial volume effect
ReLU	rectified linear unit
SC	spinal cord
SCI	spinal cord injury
SCT	spinal cord toolbox
SVM	support vector machine
VF	ventral funiculus

CHAPTER 1 INTRODUCTION

Multiple Sclerosis (MS) is a chronic immune mediated disease of the central nervous system (CNS) affecting around 100,000 people in Canada ("MS Society of Canada," n.d.). MS is probably the most frequent cause of neurological disability in young adults, with high variability in clinical expression. The pathologic hallmark of MS is the focal areas of myelin loss within the central nervous system, known as lesions (Popescu and Lucchinetti, 2012). MS clinical progress is characterised by episodes with recovery, episodes leaving persistent deficits, and secondary progression, through motor, sensorial, visual and cognitive impairment (Compston and Coles, 2002).

The spinal cord (SC) is the main bridge between the brain and the CNS, guaranteeing the conduction of both motor and sensory signals, and containing neurons associated with key functions such as locomotion (Rossignol, 2006). Conventional Magnetic Resonance Imaging (MRI) is among the recommended methods to assess SC damages, due to its non-invasiveness and sensitivity to parenchymal tissue. However, SC MRI has long been technically challenging, mainly due to the small sizes of the cord, as well as the image distortions induced by vertebral structures or respiration mechanism. Recent improvements in pulse sequences and hardware (Breckwoldt et al., 2017; Kilsdonk et al., 2016; Stroman et al., 2014; White et al., 2011) significantly enhanced the quality of SC MR images and then allowed new perspectives for SC data analyses.

During the past two decades, the analysis of SC MS lesions has raised an important interest (Hua et al., 2015; Kearney et al., 2015a, 2013; Rocca et al., 2013), given their important value for diagnosis and prognosis of MS (Arrambide et al., 2018; Sombekke et al., 2013; Thorpe et al., 1996). Furthermore, the quantification of cord atrophy, frequently developed by MS patients (Bakshi et al., 2005), has been confirmed to be clinically relevant and to correlate with clinical disability (Kearney et al., 2014; Losseff et al., 1996; Lundell et al., 2017; Rocca et al., 2013, 2011; Valsasina et al., 2018). Consequently, segmenting the SC and its lesions may provide valuable quantitative characterization of the disease. Besides MS pathology, SC segmentation could allow an objective quantitative assessment of SC morphometry (Fradet et al., 2014;

Papinutto et al., 2015) or the development of disease biomarker (Martin et al., 2017; Nakamura et al., 2008; Paquin et al., 2018).

However, manual segmentation is prone to inter- and intra-variability and is highly time consuming. Hence, there is a need for robust and automatic segmentation tools for the SC and MS lesions within it. The implementation of these methods as part of automatic pipelines could open the door to SC MRI data analyses on large multi-center cohorts, towards further understanding the spinal cord pathophysiology.

The main goal of this Master's project was to develop a method to automatically segment the MS lesions on SC MRI data (**G1**). To this end, it first required to develop automatic tools for the following intermediary tasks: (i) detect the spinal cord (**G1.1**), and then (ii) segment it (**G1.2**). Once validated against a multi-center dataset and integrated into an open-source software (De Leener et al., 2017) (**G2**), these tools were incorporated into quantitative analysis pipelines on (i) Amyotrophic Lateral Sclerosis (ALS) (**G2.1**) and, (ii) MS (**G2.2**) patients. Note that, in this thesis, the emphasis will be put on **G1**.

The 5 remaining chapters are organized as follows. **Chapter 2** overviews the background knowledge gravitating around this research project. **Chapter 3** gives a brief presentation of the research methodology, the avenues investigated to tackle its objectives and the scientific publications resulting from this project. **Chapter 4** presents the article meeting the goals *G1., G1.1 and G1.2*. **Chapter 5** is a general discussion about the performed works and suggests some potential lines of investigations for further works. **Chapter 6** draws conclusions from this Master's project.

CHAPTER 2 LITERATURE REVIEW

This chapter aims at providing the reader an overview of the background, the rational, the challenges and the related works of this Master's project. To this end, it spans: the spinal cord (**section 2.1**), the spinal cord MRI data features (**section 2.2**), the MS and ALS diseases (**section 2.3**), the automatic segmentation of the SC and MS lesions related works (**section 2.4**), and general concepts of Convolutional Neural Networks (CNN) used for the segmentation task on biomedical images (**section 2.4**).

2.1 Spinal cord

This section briefly presents the central nervous system (**subsection 2.2.1**), the spinal cord tracts (**subsection 2.2.2**), the spinal cord organization (**subsection 2.2.3**) and the spinal cord morphometry (**subsection 2.2.4**).

2.1.1 Central nervous system

The nervous system is composed of all the nerve cells present in our body. The nervous system receives information from our sensory organs, processes the information, prompts reactions and controls the metabolic processes. The nerve cells are made up a cell body, containing the nucleus, and several extensions, which are two types: (i) dendrites, (ii) axons (**Figure 2.1**). The dendrites receive signals and transfer it to the cell body, while the axons take signals away from it.

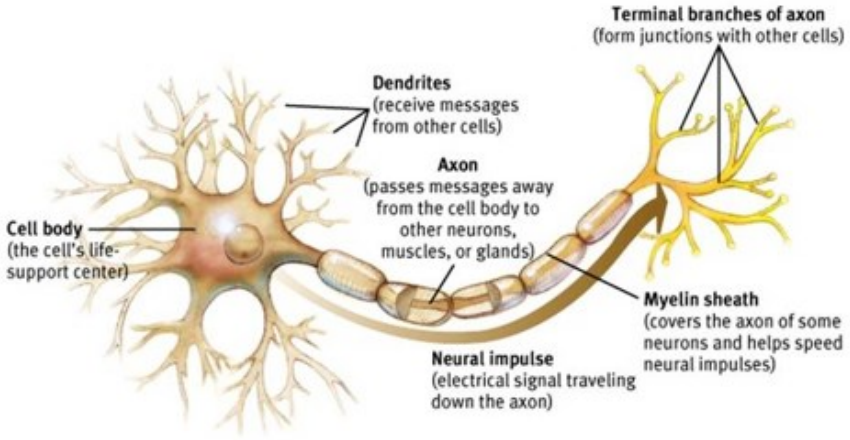


Figure 2.1: Neuron.

A neuron is composed of a cell body, dendrites and axon surrounded by myelin. This figure was extracted from http://www.appsychology.com/Book/Biological/neuroscience.htm

The nervous system is made up of two parts: (i) the CNS and (ii) the Peripheral Nervous System (PNS). The CNS interacts with the rest of the body through the PNS, which lies outside of the brain and spinal cord and is connected with the rest of the organism. In the PNS, axons or nerve fibbers conduct information from and to the CNS. The CNS incorporates sensory or motor information, coming from the whole organism and regulates the physiological activity. For instances, it controls the breathing, heart rate, hormonal secretion and body temperature. It harmonizes and organizes our thoughts, emotions, desires, movements and memories.

The CNS is composed of the association of the brain and the spinal cord. Although brain is commonly considered as the most fascinating part of the human body, we will try, in this thesis, to put the light on some attractive aspects of the spinal cord.

2.1.2 Spinal cord tracts

The spinal cord is the main bridge between the brain and the rest of the body, connected to the brain through the brain stem. Motor instructions from the brain transit from the brain through the spine to the muscles (descending tracts on **Figure 2.2**). Sensory information travel from sensory tissues through the cord towards the brain (ascending tracts on **Figure 2.2**). The spinal cord also controls some reflexive responses and hosts neurons associated with key functions such as locomotion (Rossignol, 2006). The spinal cord nerves can indeed coordinate walking muscles independently of brain activity.

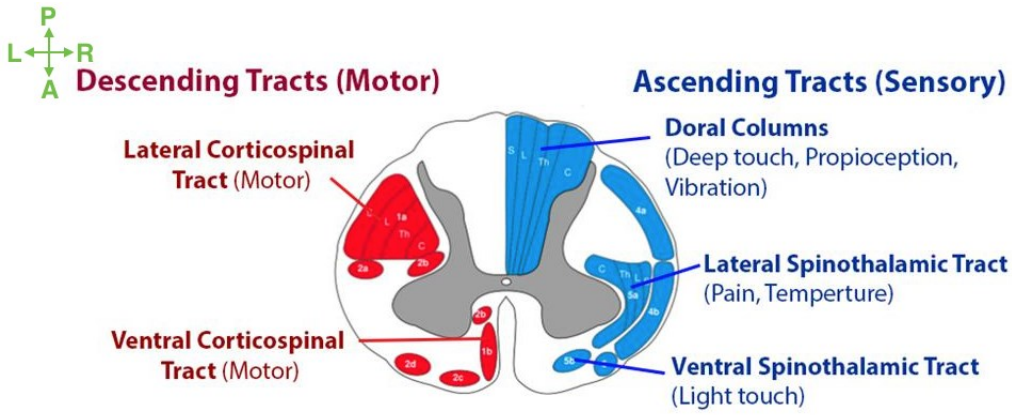


Figure 2.2: Spinal cord tracts, ascending (sensory) and descending (motor).

Each tract or column is associated with functions indicated into brackets. Figure extracted from https://commons.wikimedia.org/wiki/File:Spinal_cord_tracts_-_English.svg. P: Posterior, A: Anterior, L: Left, R: Right.

The spinal cord tissue is organized into grey and white matter. Grey matter contains the cell bodies of the neurons, while white matter contains the axons. More precisely, white matter is composed of bundles of axons, transmitting sensory or motor information, belonging either to ascending or descending tracts (Lévy et al., 2015). The white matter can be subdivided into 3 columns: (i) dorsal column (DC), (ii) lateral funiculus (LF), (iii) ventral funiculus (VF), each of them having a very specific role. For instance, DC transmit sensory signal of the touch perception to the somatosensory cortex. Regarding the grey matter and its butterfly-shape, the front "wings'' (named horns) host motor nerve cells, while the back horns contain sensory nerve cells (see **Figure 2.3**). As a result, the spatial localization of spinal cord damages is highly associated with the functional impairment, since each spinal pathway has a particular function in the CNS.

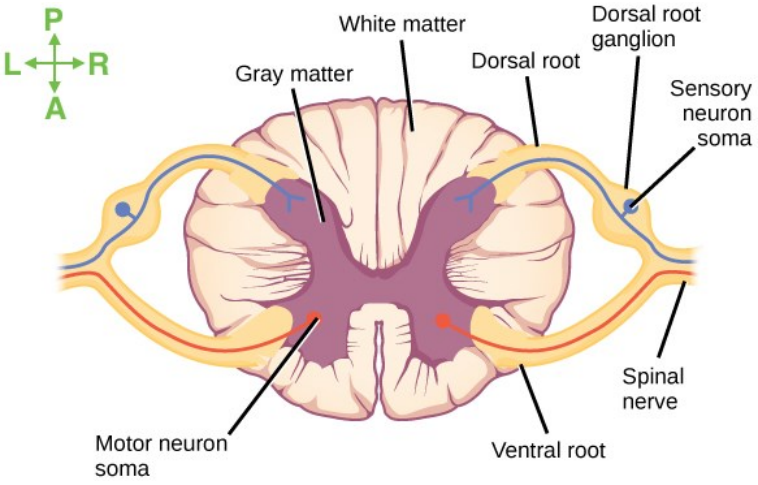


Figure 2.3: Spinal root.

Figure showing the motor nerves (red) entering in the grey matter spinal cord through the ventral root, while sensory nerve transit from the body to the spinal cord through the dorsal root. Figure extracted from http://www.newhealthadvisor.com/Spinal-Cord-Cross-Section.html. P: Posterior, A: Anterior, L: Left, R: Right.

2.1.3 Spinal canal

The spinal cord is surrounded by a layer of cerebrospinal fluid (CSF) and then protected by the vertebral column. The vertebral discs cushion the vertebral bodies and allow flexibility to the spinal cord (see **Figure 2.4 A**). The human vertebral column is commonly divided between cervical (7 levels), thoracic (12), lumbar (5) and sacral (5) sections (see **Figure 2.4 B**). Each region of the spinal cord innervates a particular group of organ, muscle or skin (see **Figure 2.4 C**). Spinal nerve roots come in the spinal cord either through the sensory or dorsal root to convey sensory information to the CNS, or through the motor or ventral root to transmit motor information to the periphery. Consequently, a damage to the spinal cord could result in a functional impairment below the level of the injury.

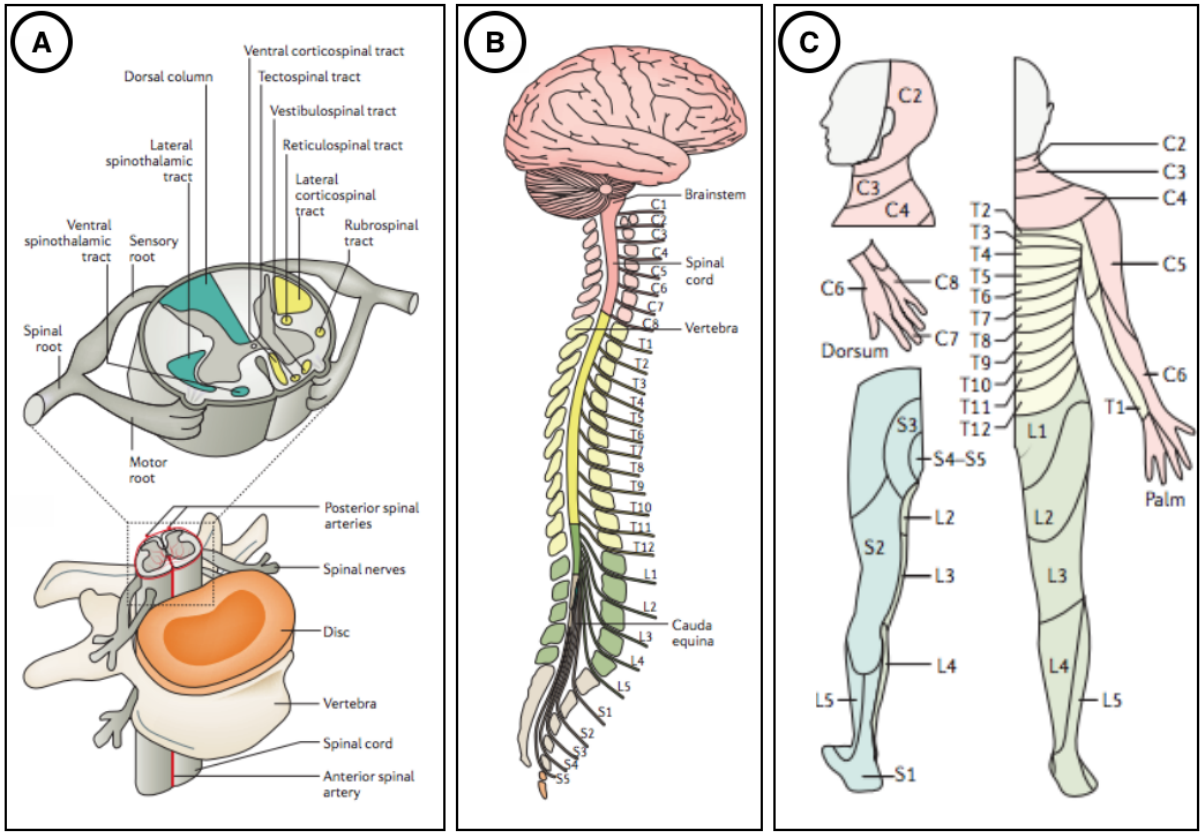


Figure 2.4: Anatomy of the spinal cord.

(A) The spinal cord is made up of grey and white matter. The white matter is subdivided into ascending (green) and descending (yellow) tracts. Spinal nerve roots come in the spinal cord through the sensory, dorsal, ventral or motor roots to transmit motor or sensory information. The vertebral column, mainly composed of discs and vertebral bodies, as well as a protective layer of cerebrospinal fluid protect the spinal cord. (B) The vertebral column is segmented into 7 cervical (red), 12 thoracic (yellow), 5 lumbar (green) and 5 sacral (grey) vertebrae. Each region of the spinal cord innervates a particular region of the skin (C). Figure adapted from (Ahuja et al., 2017).

2.1.4 Spinal cord morphometry

Spinal cord has a tubular and ellipsoidal shape, with small dimensions, which vary depending on the vertebral level. Indeed, the mean cross-sectional area of the cervical cord is about 91mm^2, compared to 68mm^2 for the lumbar cord (Fradet et al., 2014). For adult humans, the spinal cord has an average length of 45cm (Goto and Otsuka, 1997). Spinal cord curvature follows the vertebral column and differs according to the subject and its position. In particular, a

significant enlargement of the spinal cord is observed between C4 and C5 (“cervical enlargement”) vertebral levels, as well as between T12 and L1 (“lumbar enlargement”), due to expanded neural input and output to the upper and lower limbs respectively.

2.2 Spinal cord magnetic resonance imaging

MRI is among the recommended methods to assess spinal cord damages due to its non-invasiveness and sensitivity to parenchymal tissue. In particular, non-invasive imaging methods are essential to image the spinal cord since the surrounding structures make the cord inaccessible for human in-vivo research. Besides of this, the surrounding structures combined with the anatomical configuration of the spinal cord induce important challenges for MR imaging. In this section, we discuss important data features (**sub-section 2.2.1**) and main challenges of the spinal cord MR imaging (**sub-section 2.2.2**).

2.2.1 Data features

A main challenge of this project was to face the large heterogeneity of data features commonly exhibited by a multi-center dataset, such as the resolution, the field of view, orientation, and the MR contrast. **Figure 2.5** depicts some heterogeneous spinal cord MRI samples.

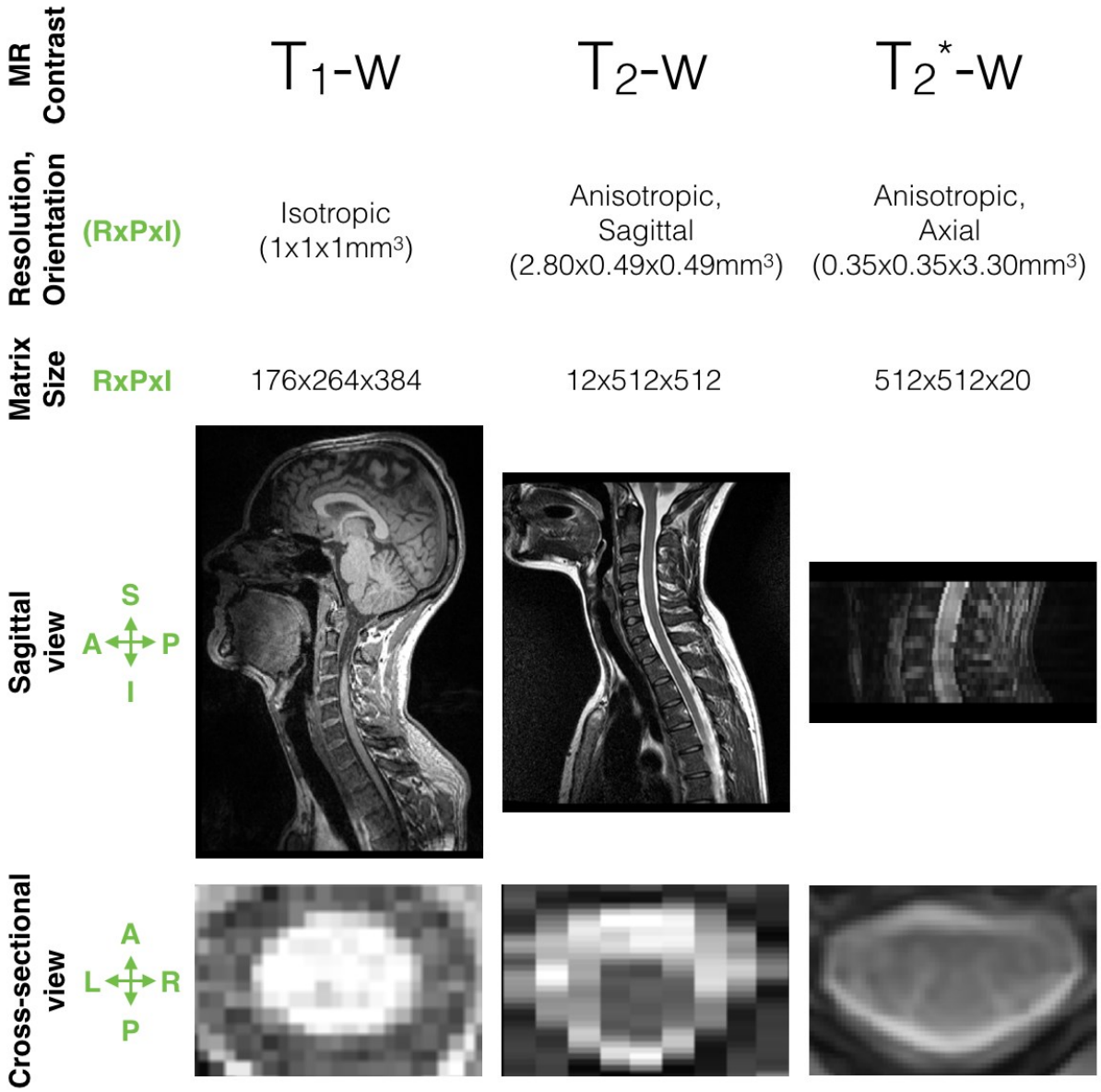


Figure 2.5: Samples of spinal MRI data.

Samples differ in terms of MR contrast, resolution, orientation, matrix size, field of view. The heterogeneity of these features are visible on the sagittal and cross-sectional view. Note that the spinal cords do not have the same dimensions on the cross-sectional view since they are not from the same vertebral level. P: Posterior, A: Anterior, L: Left, R: Right, I: Inferior, S: Superior. RxPxI: Right-to-Left x Posterior-to-Anterior x Inferior-to-Superior.

2.2.1.1 Resolution and field of view

In MRI, spatial resolution is the size of the voxels in the 3D acquired image. The resolution can be anisotropic (i.e. 2D acquisition) or isotropic (i.e. 3D acquisition) and depends

on the matrix size, the field-of-view (FOV), and the slice thickness (see second row of **Figure 2.5**). The matrix size is defined by (i) the number of frequency encoding steps in one direction, (ii) the number of phase encoding steps in the other direction of the image plane. The FOV determines the amount of coverage over which an MR image is acquired (Fourier space, k-space) or displayed (spatial space). The slice thickness is the voxel depth and is related to the maximum strength of the z-gradient coils and time restraints limiting the number of slices available. Typical resolution used in clinical routine is about 1x1 mm^2 in the image plane, and a slice thickness between 1 and 3 mm.

2.2.1.2 Orientation

Axial scans place the highest spatial resolution in the cross-sectional plane (see last row of **Figure 2.5**), favouring the spinal cord plan where the anatomy is more varied. However, the main drawback of axial scans is that they usually span a relative small Superior-to-Inferior view (see last column of **Figure 2.5**) since it is time-consuming to acquire a large number of slices.

Sagittal scans however have a worst resolution in the Left-to-Right direction but are valuable considering the small dimensions of the cord and its relative low curvature in this direction. These images allow to cover large Superior-to-Inferior view but suffer from partial volume effect (PVE) in the cross-sectional plane. Note that the partial volume effect is when several structures may contribute to the signal of the border voxels (see **Figure 2.6**). In particular, the PVE is even more important when the image slice is not orthogonal to the spinal cord axis. As a result, cross-sectional area measures, such as used for atrophy quantification, is then not recommended on sagittal scans.

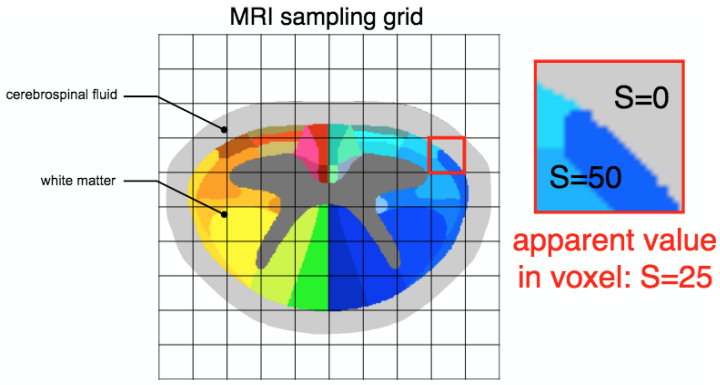


Figure 2.6: Partial volume effect in the cross-sectional plane.

The partial volume effect is when several tissue may contribute to the voxel value. For instance, both white matter and cerebrospinal fluid signals contribute to the value in voxel circled in red. Figure courtesy from Simon Lévy.

2.2.1.3 MR contrast

For this project, conventional MRI was used while new neuroimaging approaches also exist (e.g. diffusion tensor imaging, and functional MRI). Although the acquisition sequences vary largely across centers, we grouped our data in three datasets (see first row of **Figure 2.5**):

- "T_1-weighted": dark CSF / light cord, example of sequence
- "T_2-weighted": light CSF / dark cord / grey matter not visible
- "T_2*-weighted": light CSF / dark cord / grey matter visible

2.2.2 Acquisition challenges

The challenges are mainly of 3 types: (i) small cord dimensions in the cross-sectional plane, (ii) physiological motion, (iii) spatial inhomogeneous magnetic field environment.

Compared to other anatomical structures commonly imaged with MRI, like brain or heart, the spinal cord is difficult to image because of its small physical dimensions. Indeed, a typical in-plane spatial resolution of 1x1mm^2 does not allow to depict small anatomical details of the spinal cord (see last row **Figure 2.5**).

The spinal cord is surrounded by the CSF which flows, non-uniformly, back and forth in the Superior-Inferior direction induced by the heart beat (Matsuzaki et al., 1996). The CSF flow prompts the spinal cord movement inside the spinal canal, which hinder the spinal cord MR

imaging. It could cause areas of hyperintensity or signal voids. Furthermore, physiological motion induced by respiration, posture or periodic movements (e.g. heart) can make the spinal cord slightly move and causes image artifacts. Axial scans are less prone to motion artifacts (e.g. heart, lungs) thanks to the possibility to use a phase-encoding in the right/left direction.

To acquire MR images in the spinal cord is also challenged by the inhomogeneous magnetic field in this region. Different susceptibilities in the vertebrae, the intervertebral discs and the air-filled lung may produce signal loss and geometric distortions (Verma and Cohen-Adad, 2014). Axial scans are less prone to these artefacts than sagittal scans since there is less field variation across the slice thickness.

However, during the last decades, significant improvements were achieved, both in pulse sequences and hardware, in mitigating these issues (Stroman et al., 2014; Ryan Topfer et al., 2018), opening new perspectives for the spinal cord analysis.

2.3 Spinal cord damage

Spinal cord damage might cause partial or complete loss of the sensory or motor functions, that could lead to different degree of paralysis. They could result from trauma or neurodegenerative diseases. In this section, we will briefly present two diseases, the MS (**subsection 2.3.1**) and the ALS (**subsection 2.3.2**), and how they affect the spinal cord tissue.

2.3.1 Multiple sclerosis

MS is a chronic, inflammatory and demyelinating disease of the CNS. Involved by the inflammation, the myelin degeneration impairs the communication ability of the impacted areas, resulting in physical, mental or psychiatric problems. However, the disease progression (i.e. inflammation events, axonal loss and brain atrophy) are not uniform across patients. Indeed, most of the time, the inflammation disappears and reparation mechanisms (called "remyelination") could recover the myelin. Nevertheless, the inflammation and demyelination could be too high compared to the remyelination mechanism in some instances, leading to non-reversible connectivity (Compston and Coles, 2002). Depending on the disease progression over time, phenotypes of MS are defined (see **Figure 2.7**):

- *Clinically Isolated Syndrome* (CIS) patients, refers to a first episode of neurologic symptoms, caused by inflammation or demyelination.
- *Relapsing Remitting MS (RRMS)* patients (~85%) alternate over time between "spikes" of disability and (partial or full) recovery. *Secondary Progressive MS (SPMS)* patients present a constant and non-reversible progression of disability. SPMS follows an initial relapsing-remitting course.
- *Primary Progressive MS (SPMS)* patients (~15%) are characterized by a progression of the disability without early relapses or remissions. PPMS can be further characterized at different points in time as either active or not active, as well as with progression or without progression.

Note that we called *disease progression* when there is evidence of disease worsening over time ; and *relapse* an attack of new or increasing neurologic symptoms.

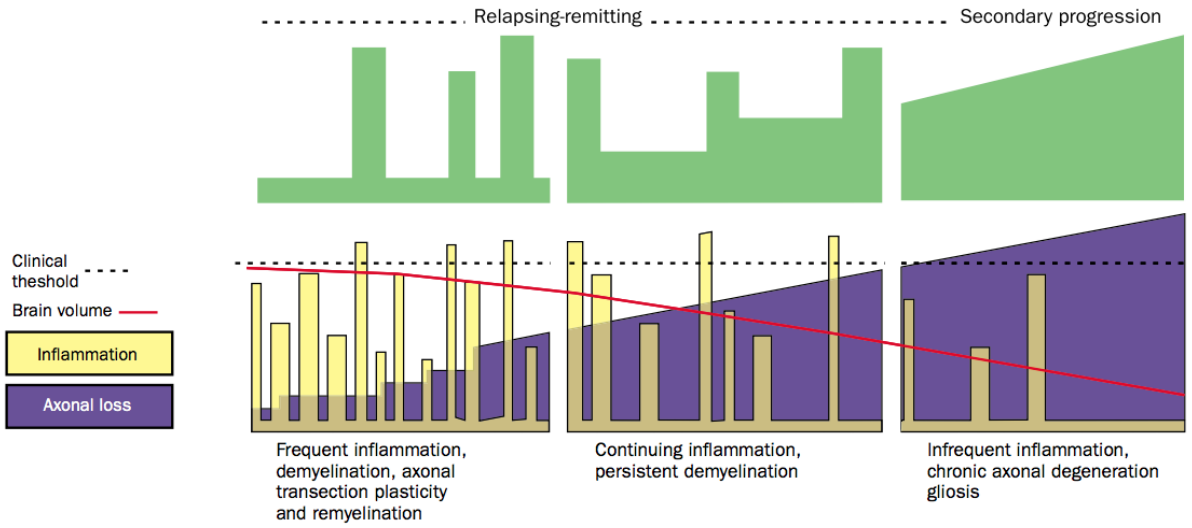


Figure 2.7: Course of the Multiple Sclerosis over time.

Cource of Multiple Scleoris over time (Compston and Coles, 2002), in terms of disability (green), inflammation (yellow), axonal loss (violet) and brain atrophy (red). The progression differs for relapsing-remitting (left) or secondary progressive (right) MS patients.

Compston et al. et al. stated that MRI presents focal or confluent abnormalities in white matter in more than 95% of patients (Compston and Coles, 2002). Demyelination occurs in plaques with various sizes and shapes, in both brain and spinal cord. **Figure 2.7** depicts some spinal cord lesions, with either a focal or diffuse appearance. The McDonald diagnostic criteria include the number and size of lesion in the brain and spinal cord detected with MRI (Polman et al., 2011). Briefly, the 2010 McDonald diagnostic criteria can be sum up as follows:

- *Dissemination in space*: At least one T_2 lesion in two of the following areas: spinal cord, periventricular area, juxtacortical area and infratentorial area
- *Dissemination in time:* A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI as compared with a baseline scan, without regard of the timing of the baseline MRI

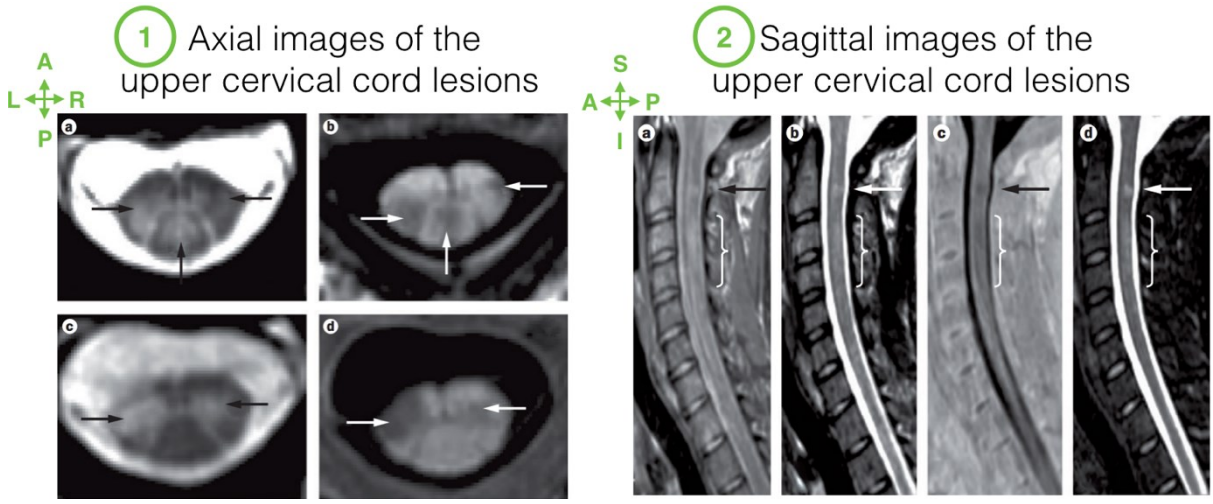


Figure 2.8: Axial and Sagittal views of cervical cord with lesions.

Axial (1) and Sagittal (2) images of the cervical cord. Arrows indicate focal lesions, visible on hyperintense (e.g. 1.a, 2.b) or hypointense abnormalities (e.g. 1.b, 2.c). Note that the lesions extent to grey matter on 1.c and 1.d axial images. Images were acquired with different sequences: fast-field echo (1.a and 1.c), phase-sensitive inversion recovery (1.c, 1.d and 2.c), proton-density weighted (2.a), T2-weighted (2.b) and short tau inversion recovery (2.d). Note that lesion detectability differs a lot across images. For instance, the diffuse abnormality (indicated by bracket) is clearly visible on (2.a) while the focal lesion is especially clear on (2.b,c,d). Figure adapted from (Kearney et al., 2015b).

In 1950, Fog observed a lesion predominance in the posterior and lateral white matter (Fog, 1950). In the last two decades, the lesion involvement in the grey matter has been shown, either through a grey-white matter extension, either confined in the grey matter (Bot et al., 2004; Lycklama et al., 2003). As explained in the section 2.1.2, the location of the white-matter lesions is correlated with their clinical impact (see **Figure 2.2**). However, the effects of grey-matter lesions are still not well understood (Kearney et al., 2013).

SC axonal loss can be vast in MS patients since Lovas et al. showed that axonal density can be reduced by 65% (Lovas et al., 2000). Spinal cord atrophy measured on MRI data has been shown to be a reliable biomarker to monitor the neurodegenerative process in MS patients (Bakshi et al., 2005). SC atrophy may be focal, around a lesion, or extended to a larger portion (Bastianello et al., 2000). It is however important to note that SC atrophy is a nonspecific measure of tissue injury since it includes axonal destruction, demyelination, and other tissue

degeneration (Bot et al., 2004). SC atrophy demonstrated moderate-to-strong correlation with disability in cross-sectional and longitudinal studies (Kearney et al., 2015a, 2014; Losseff et al., 1996; Stevenson et al., 1998). When comparing phenotypes populations, spinal cord atrophy was more pronounced in progressive forms of MS (i.e. SPMS and PPMS patients), and especially greatest in SPMS patient (Losseff et al., 1996). Among progressive forms, the rates of atrophy in the brain and in the spinal cord are different in PPMS patients, suggesting differences in the underlying pathological processes between the brain and the spinal cord (Ingle et al., 2003). Finally, cord atrophy holds an important diagnosis value since it is present already in the earliest stages of the disease (Biberacher et al., 2015).

2.3.2 Amyotrophic lateral sclerosis

ALS is a progressive and highly heterogeneous neurodegenerative disease, mainly affecting the motor neurons in the cerebral cortex, brainstem and spinal cord. The degeneration of the lower and upper motor neurons results in muscle weakness, atrophy or clumsiness. The muscle size decrease leads to loss of autonomy, speaking or breathing impairment, which can lead to death. ALS is characterized by a short survival rate (median survival from onset: 23–52 months) (Blumenfeld, n.d.; Brooks, 1996). Although implication of genetics has been recognized (Renton et al., 2014), the knowledge of the mechanisms leading to this disease are still insufficient and the cause is not known for most cases. At present, this disease remains incurable and there is a lack of proper treatment.

The large heterogeneity of clinical features present in the early course of ALS hampers an absolute diagnosis (Brooks, 1994). Briefly, the ALS diagnosis focuses on the presence of signs of upper and lower motor neuron dysfunction within the same body segments. In most cases, ALS onset is located in spinal or bulbar regions (see **Figure 2.9 A**). However, ALS clinical presentation and progression are highly heterogeneous among patients (see **Figure 2.9 B**). The assessment of the timing and the monitoring method of a patient account for an important reason for the clinical variability (Swinnen and Robberecht, 2014). The development of robust biomarkers could better categorize clinical phenotype, improve prognosis, identify the true biologic effects of drug testing in clinical trials, and more generally promote a better understanding of the physiopathology processes of the disease (Turner and Verstraete, 2015). In particular, biomarkers of the spinal cord can potentially provide a relevant measure of the

degeneration of lower motor neurons (Cohen-Adad et al., 2013; El Mendili et al., 2014; Turner and Benatar, 2015). El Mendili et al. established that cord atrophy measured with MRI was a valuable biomarker of disease progression, improving prediction of the arm-revised ALS Functional Rating Scale (ALSFRS-R) subscore at 1 year (El Mendili et al., 2014). Cohen-Adad et al. showed that local spinal cord atrophy was correlated with deficiency in the corresponding muscle territory (e.g. C4 level for deltoid and C7 level for hand muscles). The idea behind it is that detecting local SC atrophy on MRI data, corresponding to specifically altered muscles, could be a reliable non-invasive quantification of lower motor neuron degeneration.

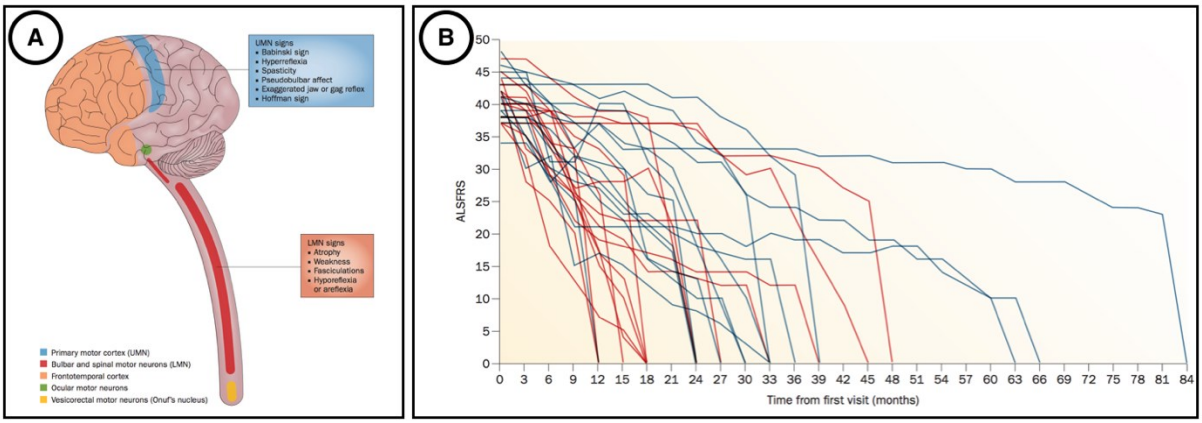


Figure 2.9: Phenotypic variability of ALS patients.

(A) shows the preferential locations of neuronal areas impaired by ALS disease. The main involved regions are the upper motor neurons (UMN) in the motor cortex (blue) and the lower motor neurons (LMN) in bulbar or spinal regions (red). (B) illustrates the variability of disease progression scored by the ALSFRS (ALS Functional Rating Scale), from the diagnostic onset to death. The 30 patients involved here had either a bulbar onset (red), either a spinal onset (blue). Figure adapted from (Swinnen and Robberecht, 2014).

2.4 Automatic segmentation of the spinal cord on MRI data

The automatic segmentation and/or detection of the spinal cord on MRI data is a necessary first step towards the segmentation of MS lesions within it. Moreover, the characterization and localization of the spinal cord impairments is clinically relevant and holds an important prognostic value for functional recovery. In this section, we present a literature review of the preprocessing methods (**sub-section 2.4.1**) frequently used before performing an automatic detection (**sub-section 2.4.2**) or segmentation (**sub-section 2.4.3**) of the spinal cord. A brief review of automatic segmentation methods used in the brain for MS lesions is also introduced (**sub-section 2.4.4**).

2.4.1 Preprocessing

Most automatic segmentation frameworks rely on a first stage of pre-processing for data preparation and/or enhancement. Overall, these preprocessing steps have to compromise the robustness across multi-center datasets, flexibility in terms of number of tunable parameters, computation time and efficiency to facilitate the subsequent segmentation.

2.4.1.1 Field of view cropping

FOV cropping or region of interest assumptions are required by some algorithms to guarantee the spinal cord to be approximately centered in the image (Koh et al., 2011, 2010; Perone et al., 2017). This cropping could certainly help the spinal cord detection and/or the class imbalance by reducing the background in the investigated volume to segment. However, an automatic cropping is not always feasible without prior knowledge or detection, especially in multi-center datasets where the FOV can substantially vary across centers. On sagittal scans commonly presenting a small Right-to-Left coverage, automatic cropping could suffer from an incorrect position of the patient in the scanner. Finally, because of the spinal cord curvature, the cord frequently spans a large portion of the Anterior-to-Posterior axis.

To overcome this problem, an automatic spinal cord detection module, prior to the segmentation, could be of interest, as done by De Leener et al. (De Leener et al., 2014).

2.4.1.2 Image denoising

Noise in an MRI image is produced by the static fluctuation of signal intensity, mainly caused by (i) the molecular movement of charged particles in the subject body, (ii) electrical resistance from some hardware components (e.g. receiver coils). Denoising is an important step to enhance image quality. However, a perilous issue in the restoration of the image is the dilemma of noise removal while preserving the relevant image information (Buades et al., 2005).

One of the most used method of denoising for 3D MRI data is certainly the one proposed by Coupe et al. (Coupe et al., 2008) based on block-wise non-local-means filters. Briefly, the non-local-means filter exploits the redundancy property of MRI images to remove the noise. Most of the previously proposed denoising approaches restore the intensity of each voxel by averaging the intensities of its neighboring voxels. In this approach, the weight involving voxels in the average is based on the intensity similarity between their neighborhoods (blue block in **Figure 2.10**) and the neighborhood of the voxel under study (red block in **Figure 2.10**). This method was confirmed to efficiently remove the noise while preserving the edges. **Figure 2.10** depicts its utilization on spinal cord data, where it could be relevant for performing automatic grey matter segmentation since it increases the grey/white matter contrast.

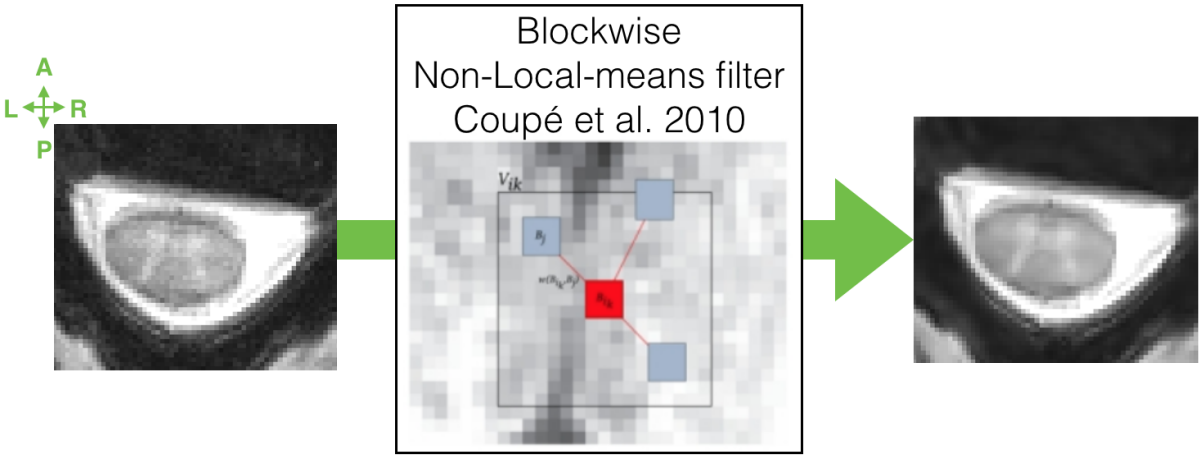


Figure 2.10: Block-wise Non-Local-means filter.

The Block-wise Non-Local-means filter (Coupe et al., 2008) is applied to the left image, and results in the right one. This algorithm is illustrated: the restored value of the block of voxels B_{ik} results from the weighted average of all the B_j blocks within the volume V_{ik}.

A critical review assessed the effects of denoising techniques (Gaussian filter, anisotropic diffusion, wavelet, and non-local-mean) on MRI brain tumor segmentation (Diaz et al., 2011).

Although the image noise was lessened by using denoising algorithms, they could introduce undesirable artifacts, harmful for the segmentation task. However, they recognized non-local-means algorithm as the most appropriated denoising technique for brain tumor segmentation.

2.4.1.3 Image normalization

One of the main difficulties with non-quantitative MR scans is that intensity values do not have a fixed meaning, not even for data acquired with the same protocol or with the same scanner, for the same imaged organ...etc (see **Figure 2.11 A**). Therefore, a direct comparison of intensities between two MR scans is not feasible since the intensities are not standardized. Image normalization is a process of mapping similar tissue types to closer intensity values (see **Figure 2.11 B**). In particular, intensity normalization has played important role for a large number of parametric supervised automatic image segmentation methods, which rely on the intensity distributions on a standardized intensity range.

Nyúl et al. proposed a method including (i) a training stage to learn a set of intensity landmarks, based on the histogram percentiles and (ii) a transformation stage to map the original intensities between two landmarks, into the corresponding learned landmarks using linear transformations (Nyúl and Udupa, 1999). Improvements of this initial version were then proposed, especially using deciles piece-wise linear mapping (Nyúl et al., 2000). This method remains widely used as pre-processing step for MRI data analysis/segmentation (Koch et al., 2017; Pereira et al., 2016; Xiao et al., 2015) and was validated on MS MRI data (Shah et al., 2011). Pereira et al. showed that CNN-based classifiers also improve after Nyúl normalization (mean gain of 4.6% in their evaluation metrics), at least in the context of brain tumor segmentation on MRI data (Pereira et al., 2016). Recently, other methods were proposed (Roy et al., 2013; Sun et al., 2015), arguing that landmark-based methods could suffer from the difficulty of learning reliable landmarks and from time-consuming hyper-parameter tuning.

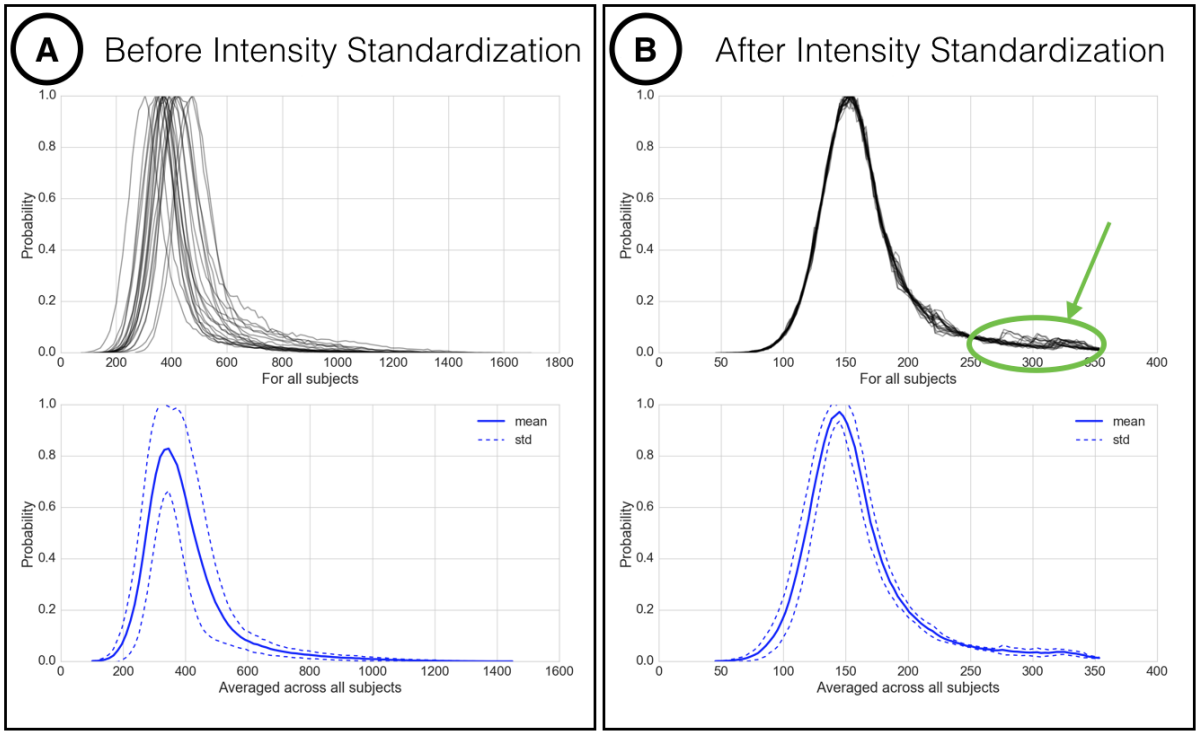


Figure 2.11: Intensity normalization proposed by Nyúl et al.

Illustration of the intensity normalization proposed by Nyúl et al. (Nyúl and Udupa, 1999) on 20 spinal cord MRI volumes of MS patients, acquired in the same center with the same sequence. The top row histograms represent the intensity distribution computed within the spinal cord of each patient, while the bottom row shows the mean (and standard deviation) histograms across patients. Spinal cord intensities across subjects have closer values after standardization (B). Moreover, some outliers (circled in green) are clearly more visible than before standardization (A), which could help the detectability of MS lesions.

2.4.1.4 Bias field correction

In order to compensate for the effect of magnetic field inhomogeneities during image acquisition, bias field corrections methods are commonly employed (Vovk et al., 2007). Bias field correction is notably relevant for spinal cord images where intensity variation induced by the B_1+ and B_1- inhomogeneities are frequent (see **section 2.2.2**). N3 and N4 algorithms are among the most prevalent techniques used in MRI (Sled et al., 1998; Tustison et al., 2010). Briefly, these methods perform a non-parametric, non-uniform, intensity normalization by estimating the multiplicative bias field among true tissue intensities, in order to correct it using B-spline least-squares fitting. Although N4 correction provides satisfying results (see **Figure 2.12**),

to find appropriated hyper-parameters is difficult and time-consuming. Moreover, issues are often noticed at the extreme slices of the data since they are related to the extreme points of the regularized mesh.

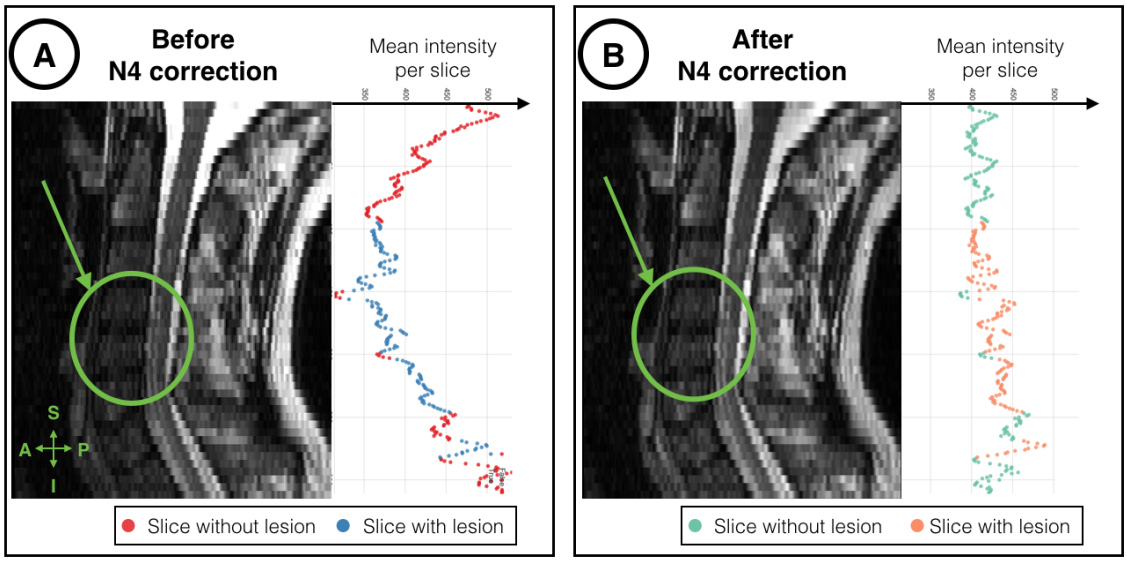


Figure 2.12: Bias field correction using N4ITK algorithm.

Comparison of a sagittal view of a spinal cord MRI data before (A) and after (B) bias field correction using N4ITK algorithm (Tustison et al., 2010). The field inhomogeneity (circled in green) has been partially corrected since the variation of the mean intensity per slice (along the Superior-to-Inferior axis) has been reduced.

2.4.2 Automatic detection of the spinal cord

2.4.2.1 Rational and challenges

Detecting the spinal cord position within MRI volumes is an essential step for automating quantitative analysis pipelines. For instance, this step plays a key role in recent published pipelines such as spinal cord (De Leener et al., 2014; Horsfield et al., 2010) and grey matter (Dupont et al., 2017; Prados et al., 2017) segmentations, template registration (De Leener et al., 2018; Stroman et al., 2008) and B0 susceptibility-related distortion correction (R. Topfer et al., 2018; Vannesjo et al., 2017).

While detecting the spinal cord might appear as an elementary computer vision task, it is much more challenging to achieve it accurately and robustly across a large variety of spinal cord shapes (see **Figure 2.13**), craniocaudal vertebral length, pathologies (e.g. compressed or

atrophied cords or hyperintense lesion areas), image FOV, image resolution and orientation, types of contrast and image artifacts (e.g. susceptibility, motion, chemical shift, ghosting, blurring, Gibbs). In the particular case of spinal cord segmentation, a recent review paper identified 20 recently-published spinal cord segmentation methods requiring manual intervention for initialization (De Leener et al., 2016). Especially, a number of these methods require the user to manually identify specific anatomic landmarks within the spinal cord. Hence, these methods could potentially be made fully automatic if initialized with a robust automated spinal cord centerline detection module.

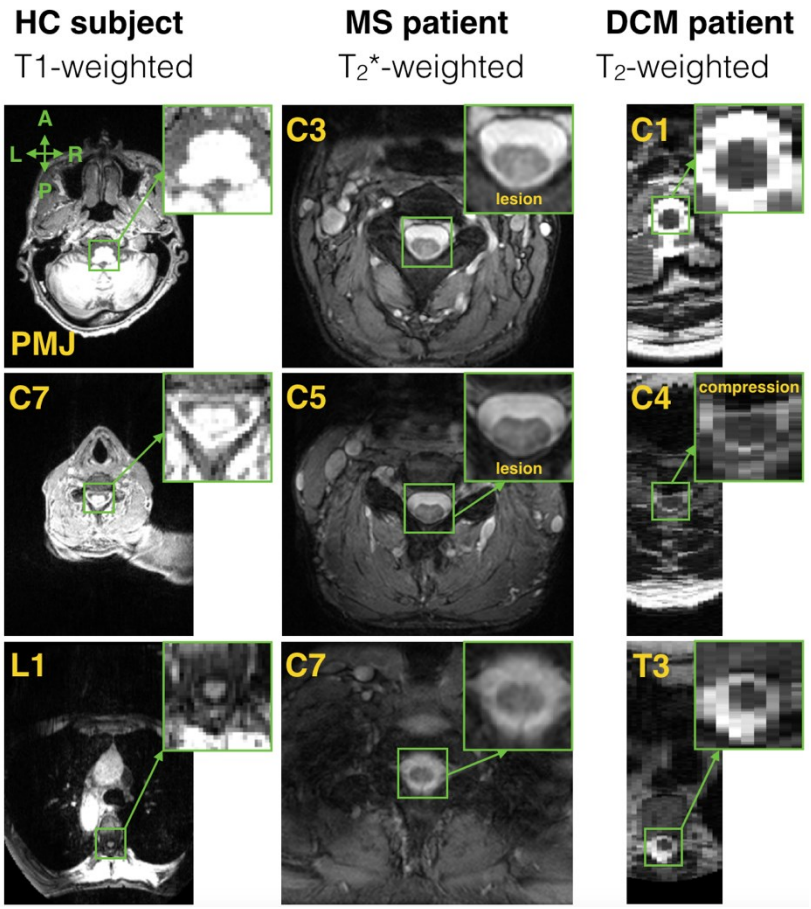


Figure 2.13: Samples of spinal cord MRI data in the cross-sectional plane.

This figure shows a large heterogeneity in terms of cord appearance depending on the MR contrast (T_1-, T_2*-, T_2-weighted), the vertebral levels (yellow, C: cervical, T: thoracic, L: lumbar, PMJ: pontomedullary junction), the resolution in the cross-sectional plane, the pathologies (compression, lesion). DCM: Degenerative cervical myelopathy, HC: healthy control, MS: Multiple sclerosis.

2.4.2.2 Related works

A common approach to automatically detect the spinal cord is to target the characteristic ellipsoid shape of the spinal cord in the cross-sectional plane. De Leener et al. proposed to detect ellipsoid shape resorting to the Hough transform combined with vesselness filtering (De Leener et al., 2014). While this approach has shown good performances, ellipsoid pattern detection is challenged by severe cord atrophy with poor CSF/SC contrast or by the important partial volume effect in the cross-sectional plane of sagittal scans. Koh et al. presented an approach based on gradient vector flow field to find the candidate structures combined with a connected component analysis (Koh et al., 2010). However, this method assumes a specific region of interest, and is designed to only work on 2D images, sagittal view, T_2-weighted MR contrast. Carbonell-Caballero et al. employed a template-based method by registering a cord pattern to the subjects' scan using cross-correlation (Carbonell-Caballero et al., 2006). Although this method was extensively validated, a major drawback is its restriction to 2D T2-w axial slices of the cervical spine. Chen et al. resorted to a deformable atlas associated with topology constraints, while making no assumptions on image resolution or FOV (Chen et al., 2013). In particular, topology constraints (classification atlas of the spinal cord, CSF and background) are used to ensure the integrity of the detected spinal cord.

2.4.3 Automatic segmentation of the spinal cord

2.4.3.1 Rational

Spinal cord segmentation is a mandatory step towards automated interpretation of morphometric and multi-parametric spinal cord MRI data. For instance, we present below 2 recent studies using the spinal cord segmentation for 2 distinct applications:

- Hori et al. evaluated the sensibility of quantitative metrics as potential clinical biomarkers for degenerative cervical myelopathy patients (Hori et al., 2018). These metrics were computed from the atlas-based regions combining information from diffusion and magnetization transfer weighted images. To do so, they used spinal cord segmentation for co-registrating all available data to a common coordinate space, allowing atlas-based quantitative multi-parametric analyses.
- Ventura et al. aimed at characterizing the atrophy process of neuromyelitis optica spectrum disorder in light of the disease evolution (Ventura et al., 2016). The cross-sectional area of the

spinal cord, obtained from its segmentation, provided them a quantitative assessment of the cord atrophy.

Automatic segmentation methods have great potential for studies involving a large number of subjects or for longitudinal analyses. Indeed, manual segmentation is highly time-consuming and suffers from inter- and intra-variability.

2.4.3.2 Challenges

In this section, we summarized the main challenges of the automatic spinal cord segmentation on MRI data, even if most of them were already mentioned in the previous sections.

Although the spinal cord shape is roughly similar to an ellipsoid in the cross-sectional plane, it is not consistent across subjects and even within a subject. Papinutto et al. showed that age and gender had a significant impact on the total cross-sectional area: women having smaller values than men, elderly smaller values than younger (Papinutto et al., 2015). Within a subject, the ratio of the first and second radii can change along the Superior-to-Inferior axis (see **Figure 2.13**).

In general, the spinal cord exhibits a substantial intensity gradient with the surrounding CSF, which could be highly useful for the automation of the spinal cord delineation. However, this might be challenged by: (i) the natural cord curvature leading to portions where the spinal cord directly touches the spinal canal, (ii) the partial volume effects (cf **Figure 2.6**), (iii) pathological hyperintense signal (e.g. MS lesions) or cord deformation (e.g. trauma, ALS, SCI…), as illustrated in **Figure 2.13**.

Finally, the automatic spinal cord segmentation can be hindered by MR artefacts, as detailed in **section 2.2.2**, such as (i) CSF flow artifacts resulting in areas of hyperintensity or signal voids, (ii) Gibbs artifacts leading to ringing lines parallel to large intensity gradients in the image, (iii) static noise, (iv) signal loss due to B1- inhomogeneities, (v) contrast alterations due to B1+ inhomogeneities.

2.4.3.3 Related works

In 2014, De Leener et al. made a review of the existing semi-automatic or automatic methods for spinal cord segmentation on MRI data (De Leener et al., 2016). The authors classified spinal cord segmentation methods in the following categories:

- *Intensity-based*, examples: feature extraction (e.g. Hough transform), thresholding (e.g. Otsu), edge detection (e.g. Kalman)
- *Surface-based*, examples: active-contours (B-spline), deformable models (tubular surface propagation)
- *Image-based*, examples: template/atlas, clustering/classifiers

Among the 26 methods listed in this review paper, only 6 were fully automatized (Koh et al. 2010; De Leener et al. 2014; Carbonell-Caballero et al. 2006; Chen et al. 2013; Pezold et al. 2015; Tang et al. 2013). In most of the 20 semi-automatic methods, the required manual intervention consists in adding some landmarks within the spinal cord (as explained in the **section 2.4.2.1**). Regarding the 6 fully automatic methods, some assumptions or limitations hamper their utilization in clinical studies:

- *Region of interest assumption*, which limits its utilization for multi-center studies where the field of view is not necessary common across centers
- *MR contrast restriction*, which limits its utilization for studies with several modalities
- *2D data restriction*, while most of MRI data are 3D
- *Lack of validation on pathological cases*, which are frequently substantially more challenging than healthy subjects (e.g. compressed cord, hyperintense lesions)
- *Lack of validation across a variety of data features*, which is a required validation step before considering its utilization in multi-center studies
- *Repeatability limitation*, since most of the methods are not publicly available, which hinder its wide dissemination through research groups and clinical routine

Among the previously published methods, PropSeg (De Leener et al., 2014) received sustained attention from the community. **Figure 2.14** presents the gist of this automatic method.

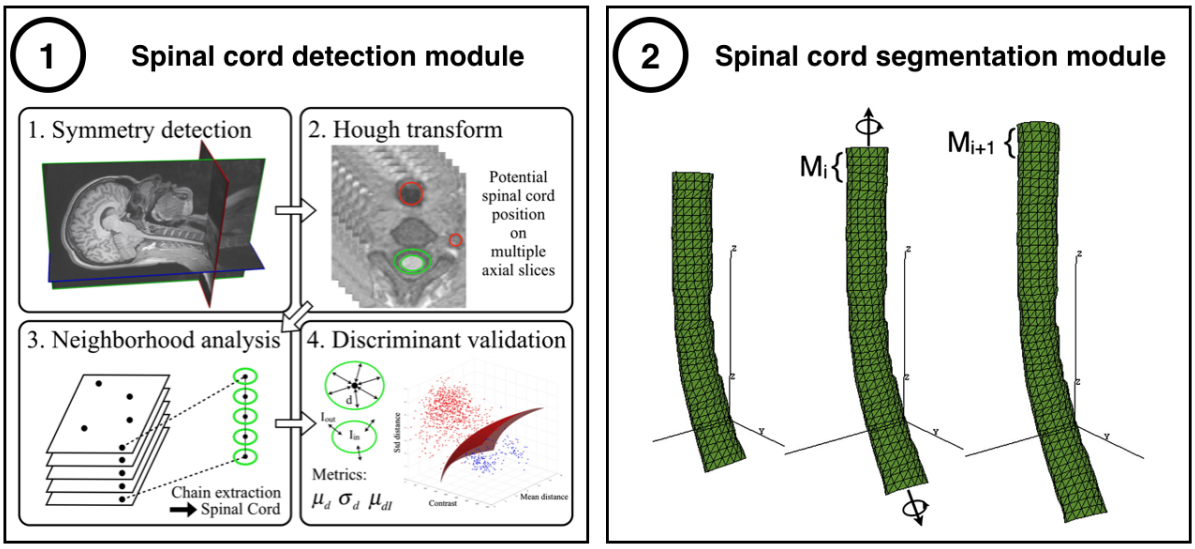


Figure 2.14: PropSeg' method for automatically segmenting the spinal cord.

The segmentation workflow begins with a spinal cord detection module (1): an elliptical Hough transform is applied around the median cross-sectional plane to find potential spinal cord position, then corrected thanks to a neighborhood analysis and a discriminant validation. The spinal cord segmentation is then performed (2) by propagation of a deformable model composed of triangular meshes (Mi). Figure adapted from (De Leener et al., 2014).

PropSeg has many advantages, which might be valuable to mention. The authors devised a model specifically designed for the spinal cord, based on anatomical priors. For instance, the method takes advantage of the tubular shape of the spinal cord thanks to a deformable model combined with a global deformation, which guarantee the continuity and integrity of the output segmentation. Moreover, PropSeg is robust to different MR contrast (T_1-, T_2, T_2*-weighted) and makes no assumption about field of view, image resolution or orientation. This is an unsupervised method, which means there is no need for manual annotation to train a model. In case of segmentation failure with the default parameters, the model can be adjusted (e.g. radius of the SC, conditions on the deformation) in order to refine the segmentation. In volumes that include brain sections, the method accurately separates brain and spine regions using a criterion on the mesh radius to stop its propagation. Furthermore, the method is fast (less than 1 minute on a typical T_2-weighted scan) and has been validated against MS patients with atrophy (Yiannakas et al., 2016). Last but not least, the framework is open-source, as part of the publicly available

software, the Spinal Cord Toolbox (De Leener et al., 2017), and further improved thanks to a massive feedback received from the community (De Leener et al., 2015).

However, it is important to note some drawbacks of this existing solution. First, the Hough transform can be challenged in instances of deformed cords, which might be critical since the segmentation performance highly depends on the detection status. As user of this method, I also noticed that (i) under-segmentation is quite frequent for T_2*-weighted images when the contrast CSF/cord is low, (ii) the segmentation tends to leak out of the cord in instances of severe cord compression.

2.4.4 Automatic segmentation of MS lesions

2.4.4.1 Rational

Several studies show evidence of the importance of SC lesions for diagnosis and prognosis of MS at clinical presentation (Arrambide et al., 2018; Sombekke et al., 2013; Thorpe et al., 1996). Arrambide et al. stated that MR scans to image spinal lesions are valuable for predicting the disease evolution of CIS patients (Arrambide et al., 2018). Sombekke et al. found that the presence of spinal cord lesions promotes diagnosing MS (Sombekke et al., 2013), in particular in CIS patients who do not fulfill brain MRI criteria but with onset of symptoms originating from the brain.

To illustrate the utilization of MS lesion segmentation in the spinal cord, we have listed below 2 recent studies using it:

- Hua et al. aimed at evaluating the relation between the cervical spinal cord lesions and the presence of thoracic cord lesions (Hua et al., 2015). They determined that the cervical spinal cord lesion involvement could predict the thoracic lesion involvement, independently of the brain involvement or the clinical status. Based on the lesion count, **Figure 2.15** (1) presents the spinal cord lesion distribution by vertebral level impacted.
- Valsasina et al. analyzed the spatial distribution of T_1-weighted cervical cord lesions in MS patients (Valsasina et al., 2018). They found that the lesion involvement was greater in the upper than lower cervical level, in the posterior than anterior cord. They also studied correlations between MS spinal cord damages and clinical disability.

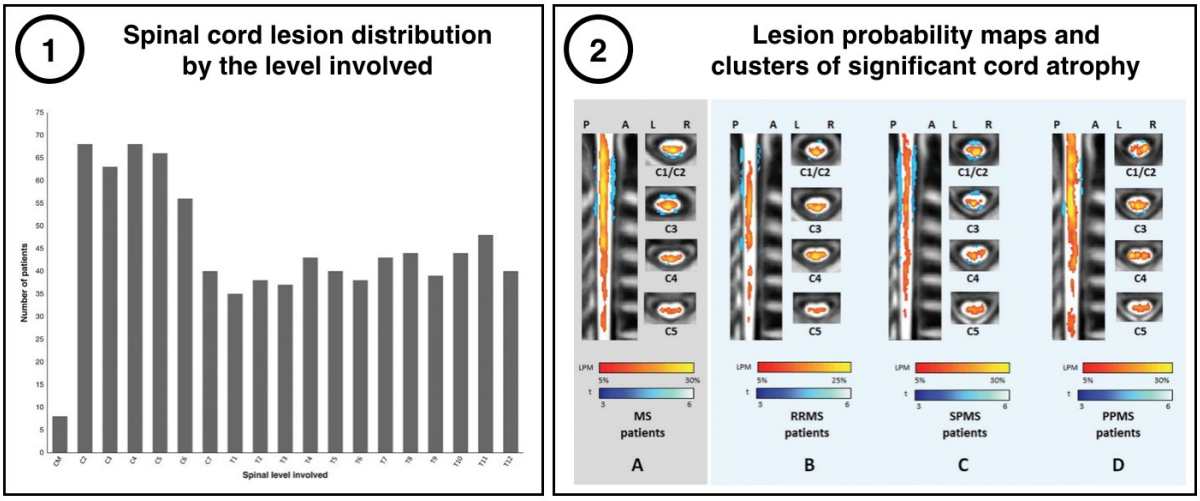


Figure 2.15: Examples of 2 clinical studies using spinal cord MS lesions segmentation.

(1) shows the spinal cord lesion distribution depending on the vertebral level (Hua et al., 2015). We observed that the lesional frequency of cervical (C1-C7) spinal levels is higher than thoracic (T1-T12) ones. (2) depicts T_1-weighted MS lesion probability maps (Red-Yellow), overlapped by significant cord atrophy clusters (Blues) when compared to healthy controls, depending of MS phenotype populations (Valsasina et al., 2018).

2.4.4.2 Challenges

The previous subsection illustrated the great interest in segmenting lesions within the spinal cord. To perform an automatic segmentation of lesions is however highly challenging. In addition to the challenges commonly faced when segmenting spinal cord MRI data, some are specific to the lesions. An intensity bias field in the Superior-to-Inferior axis (see **Figure 2.12**) could lead to a large intensity overlap between lesion and with spinal cord intensities. Moreover, the partial volume effect and the confounding intensities of lesions with normal structure (e.g. grey matter on T_2*-weighted images) may mislead the segmentation. While the spinal cord exhibits most of the time a consistent appearance, lesions are highly heterogeneous in terms of location, size and shape. Finally, lesions are small and sparse objects leading to a very small proportion of lesion voxels compared to the rest of the volume. **Figure 2.16** depicts some of these aspects.

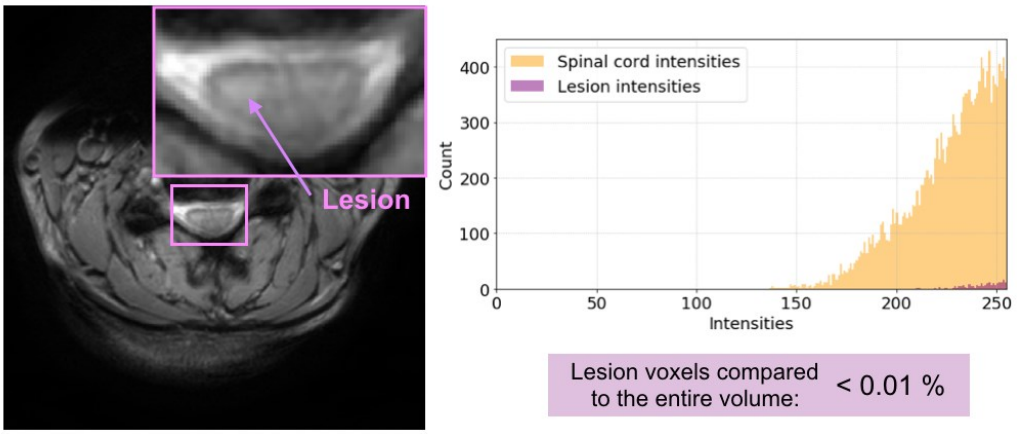


Figure 2.16: Intramedullary MS lesion – Segmentation challenges.

Sample of a spinal cord cross-sectional slice of MRI volume (T_2*-weighted contrast). A MS lesion is visible within the spinal cord (pink arrow). Lesion voxels represents less than 0.01% of the entire volume. The intensity histograms of the spinal cord and lesion voxels are shown: a large overlap between the two tissues. This is especially due to the healthy grey matter exhibiting similar intensities than lesions.

2.4.4.3 Related works

At the writing time of this section, a thorough search of the relevant literature did not yield available related work. In contrast, automatic segmentation of brain MS lesions was largely investigated in the brain (García-Lorenzo et al., 2013; Lladó et al., 2012). For instance, in 2013, Garcia-Lorenzo et al. identified 47 automatic different methods applied to brain MS lesion segmentation (García-Lorenzo et al., 2013). In this subsection, we will review some recent approaches proposed for brain lesions.

As suggested by two recently-published review papers (García-Lorenzo et al., 2013; Lladó et al., 2012), the methods of automatic MS lesions segmentation on MRI data can be classified between the supervised and unsupervised methods. Supervised methods learn lesions' characteristics based on manually labeled training data, which are used to optimize a function from inputs to the related labels. Contrary, unsupervised methods learn lesion properties by inferring a function to describe the structure or relationships between the "unlabeled" images.

Unsupervised approaches try to encode the clinical knowledge of MS disease to perform automatic segmentation on each image independently. However, the translation from the expert knowledge to the algorithm code is highly complex, mainly because of the high heterogeneity of

the disease. Most of unsupervised segmentation strategies use clustering methods, like Expectation-Maximization (Souplet et al., 2008; Van Leemput et al., 2001) or Fuzzy C-means (Shiee et al., 2010) algorithms. A common strategy consists in (i) formalize the healthy tissues properties, (ii) model them through statistical distributions (iii) and segment lesions as outliers of normal tissues distributions. These approaches highly depend on the accuracy of the normal tissues segmentation, which remains a challenging task. Another approach is to encode lesions localization using anatomical prior knowledge (e.g. anatomical atlases) to constrain the classification and then segment lesions as an additional class.

Supervised methods learn from labeled data to devise a segmentation model able to discriminate lesion and not-lesion voxels. However, these approaches usually require a large amount of data for the training to cover as many cases as possible, which is limited by (i) the high heterogeneity of lesion appearance, which requires to create a database spanning a large range of possible lesions and acquisition parameters, (iii) the difficulty of obtaining a large number of data annotations, given that manual segmentation is time-consuming. A traditional workflow of supervised approaches is: (i) data normalization in order to ensure similar intensity range between training and testing images, (ii) feature selection to discriminate lesions to normal tissues, (iii) classifier training and testing. Voxel intensity is the most common feature used by traditional classifiers, but is highly limited by challenges listed in section 2.4.4.2. Although more complex features were proposed to face these issues, feature engineering generalizes poorly to multi-center datasets. To avoid this, some approaches recently proposed to perform an efficient feature selection using PCA strategies (Roura et al., 2015) or, more notably, to use Deep Learning techniques (Brosch et al., 2016).

To summarize, supervised approaches are highly promising as long as exhaustive training is made, while unsupervised strategies are not based on ground truth and can be more flexible to new data features (e.g. MRI sequences) (Styner et al., 2008).

Despite the important efforts deployed to improve automatic segmentation methods of MS brain lesions, it still remains an open-issue. In the remainder of this subsection, we review 3 approaches published very recently and put the light on their key contributions:

- Kamnitsas et al. used multi-scale 3D Convolutional Neural Networks for segmenting lesions (Kamnitsas et al., 2017). To deal with the high class imbalance, they automatically adapt it

depending on the current batch of patches. They devised a dual pathway architecture to both local and larger contextual information (see **Figure 2.17**). Finally, they used 3D fully connected Conditional Random Field (Krähenbühl and Koltun, 2011), as post-processing step to remove false positive detections. This approach belongs to the multi-stream architectures, which aimed at combining different source of information, here a down-scaled representation in addition to high resolution local information.

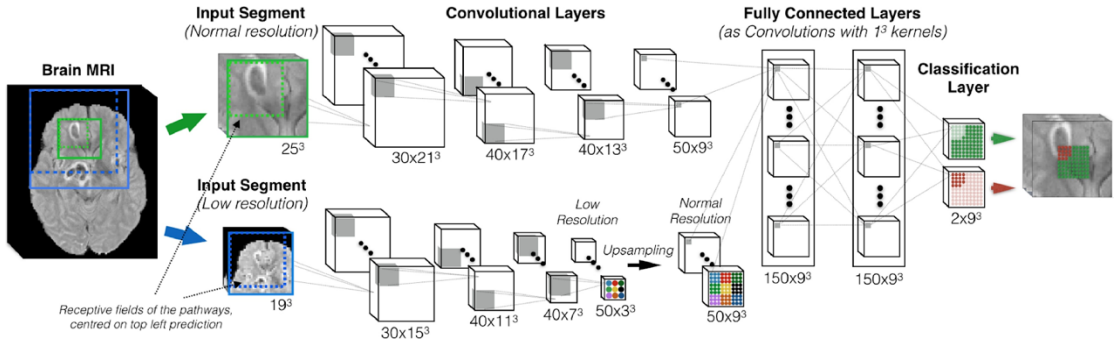
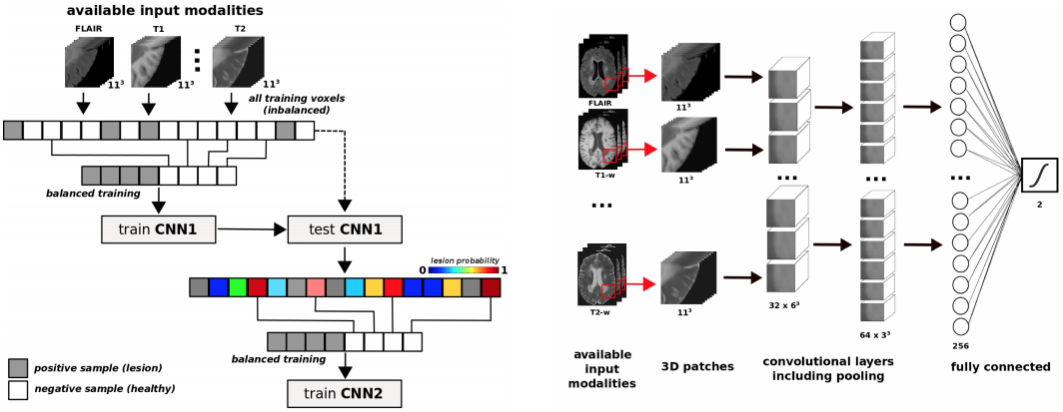


Figure 2.17: Multi-scale 3D Convolutional Neural Networks for segmenting lesions.

The input image (left) is processed at multiple scales: the local appearance of structures is learnt by the top pathway (green), while higher level features (such as the location within the brain) are captured by the bottom pathway (blue). The information coming from different scales are combined through the fully connected layers, prior to the classification layer. Figure extracted from (Kamnitsas et al., 2017) showing the multi-scale dual-pathway architecture used in this paper.

- Valverde et al. designed a cascade of two 3D CNNs (Valverde et al., 2017) (see **Figure 2.18 a.**). The general idea behind the cascade approach is to reduce the number of misclassified voxels coming from the first network by a second network trained on false positives patches of the first network. Note that, to deal with class imbalance, they under-sample the negative class within the training patches (i.e. the number of patches containing lesion voxels and patches that do not contain lesion are equal). Another key contribution of this paper is that they benefit from information coming from different image sequences, but without being restricted to (i.e. one or more image sequence can be missing).



(a) Cascade based pipeline (b) Proposed CNN architecture

Figure 2.18: Cascade of 3D CNN for segmenting lesions.

Cascade based pipeline (left) and 3D CNN architecture proposed by (Valverde et al., 2017). (a) presents the cascade pipeline where CNN1 aimed at discriminating patches containing or not lesion voxels, while CNN reduce the number of misclassified voxels on patches selected by CNN1. (b) shows the CNN architecture used by CNN1 and CNN2, where each input channel is fed by an available image sequence, without being restricted to. Figure extracted from (Valverde et al., 2017).

2.5 Convolutional Neural Networks for segmentation on MRI data

During the last few years, an outstanding interest in convolutional neural networks (CNNs) for image segmentation task has been observed, with striking performance in different areas, including biomedical images (Litjens et al., 2017a). Litjens et al. recently published a review paper based on the identification of more than 300 articles resorting to Deep Learning techniques in medical image analysis, as illustrated in **Figure 2.19**. Among the broad diversity Deep Learning approaches, they found that CNNs were more prevalent in medical imaging analysis. In this section, we will try to give a brief overview of the main concepts of the CNNs (**section 2.5.1**), and their training process (**section 2.5.2**), the key advantages of CNNs when applied to segmentation of biomedical images (**section 2.5.3**) and finally describe how the high class imbalance issue could be overcome when using CNNs (**section 2.5.4**).

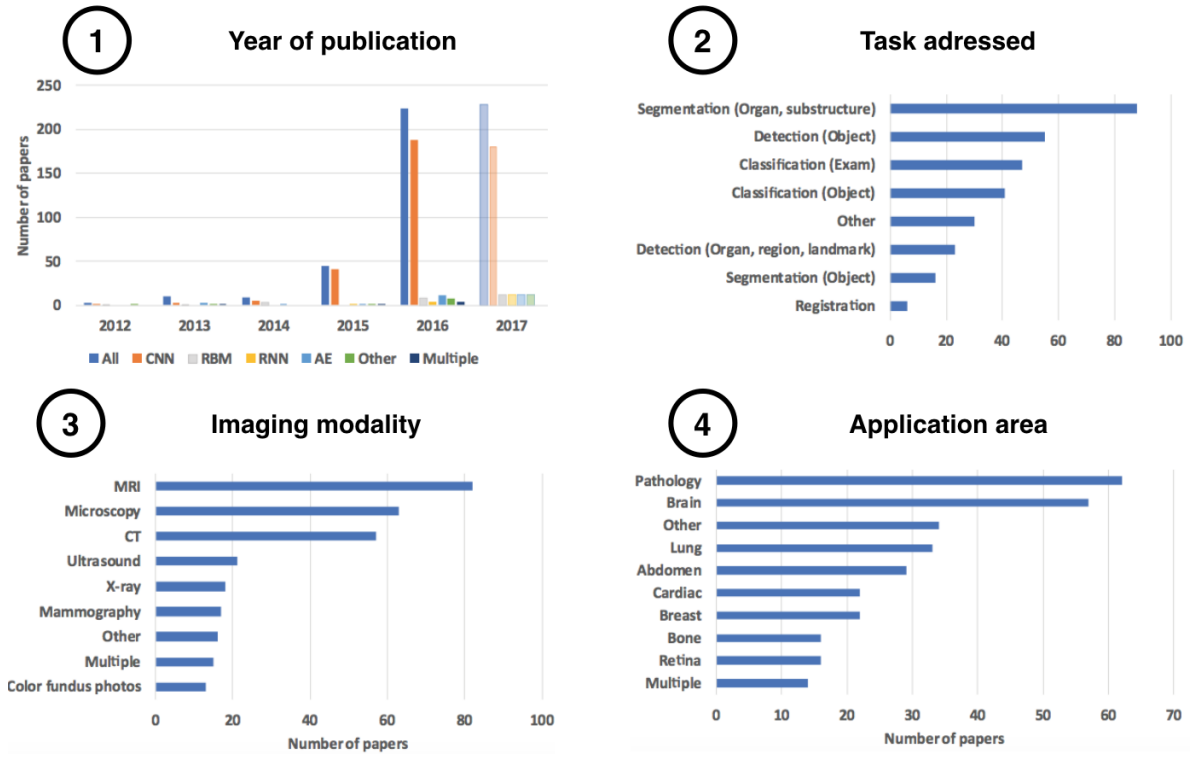


Figure 2.19: Deep Learning techniques for medical image analysis.

(1) shows an outstanding increase of studies published using Deep Learning since 2015. In particular, convolutional neural network (CNN) is the most widely employed approach. (2) shows that deep learning is mainly used for segmentation tasks (n~90), while it is still not widely used for registration (n~10). Finally, most papers employing deep learning techniques applied them to MRI data (3) for pathology and brain areas. Figure adapted from (Litjens et al., 2017b).

2.5.1 CNN architecture

Neural network is a machine learning technique, inspired from the brain structure (i.e. axon, synapse, dendrite, cell body, axon...etc.), where the neurons learn how to convert input signals (e.g. picture of radish) into corresponding output signals (e.g. the label "radish"). Simply put, during a training step, we "teach" the network a recognition task thanks to labeled samples (i.e. the question and the solution, e.g. the image and its label), by an optimization process (i.e. adjustment of the parameters of the network to minimize a cost function). Hence, the more labelled images the neurons are exposed to during the training step, the better it learns how to recognize other unlabelled images.

Computer algorithms extract features from the data and determine the optimal decision based on them. Traditional classification algorithms (e.g. Support Vector Machine, Random Forest) resorted to hand-crafted features, Histogram of Oriented Gradient (Dalal and Triggs, 2005) or Speeded-Up Robust Features (Bay et al., 2008/6). Feature engineering transforms the measured world into a set of characteristics whose properties are well-adapted to what algorithms can handle. If the features are not well-adapted to your goals, your model may have difficulties in making any useful predictions. Moreover, this designing is time-consuming and finding a set of generic feature hyper-parameters that works for every dataset is challenging. The game-changing advantage of CNNs is its "automated" feature engineering to find optimal features and handle indirect relationships between features and goals. In other words, CNNs find appropriate filters by a hierarchical representation learning strategy. Indeed, the features learned in first layers come together and make abstract shapes, which often have meaning in their deep layers. But, deep learning is not a magic bullet, the "problem" moves from the craft of feature design to the craft of feature strategy architecture.

To summarize, CNNs learn filters, which are not necessarily known, to reduce the amount of cost function. They try to find appropriate filters by their own, without being limited to discovered hand-crafted filters. In deeper layers, the features learned in previous layers come together and make shapes which often have meaning.

Figure 2.20 illustrates a simple 1D graph representation of CNN. Considering the simplest case, the input layer (green nodes) represents all pixels of the input image, while the output layer represents the filter responses (first column of grey nodes). Each node of the output

layer is connected to a pixel and its neighbourhood in the input layer (weighted connections with same color). The key of the CNNs is that the weights of the connections are shared, which means that they are the same for different pixels in the image while different depending on the position relative to the pixel of interest (i.e. the convolved filter is centered on this pixel). It results in an efficient learning of filters suited for the current problem.

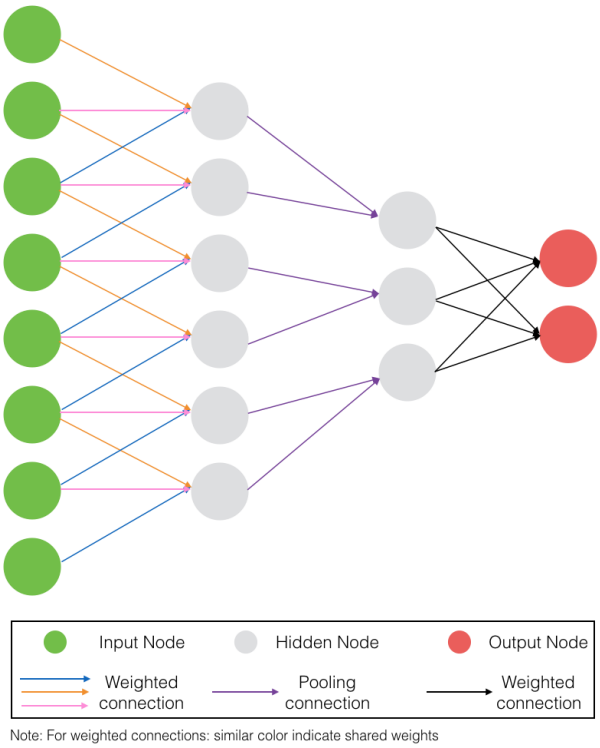


Figure 2.20: 1D representation of convolutional neural network architecture.

Each node of this graph represents a neuron. Weighted connections are shared between neighbors input nodes. The two output nodes would represent the two classes of a classification task. Figure inspired from (Litjens et al., 2017b).

Figure 2.21 presents a typical simple CNN architecture, here to classify input image between "X" and "O" classes. CNN is composed of different layers, which will be presented in the remaining of this subsection.

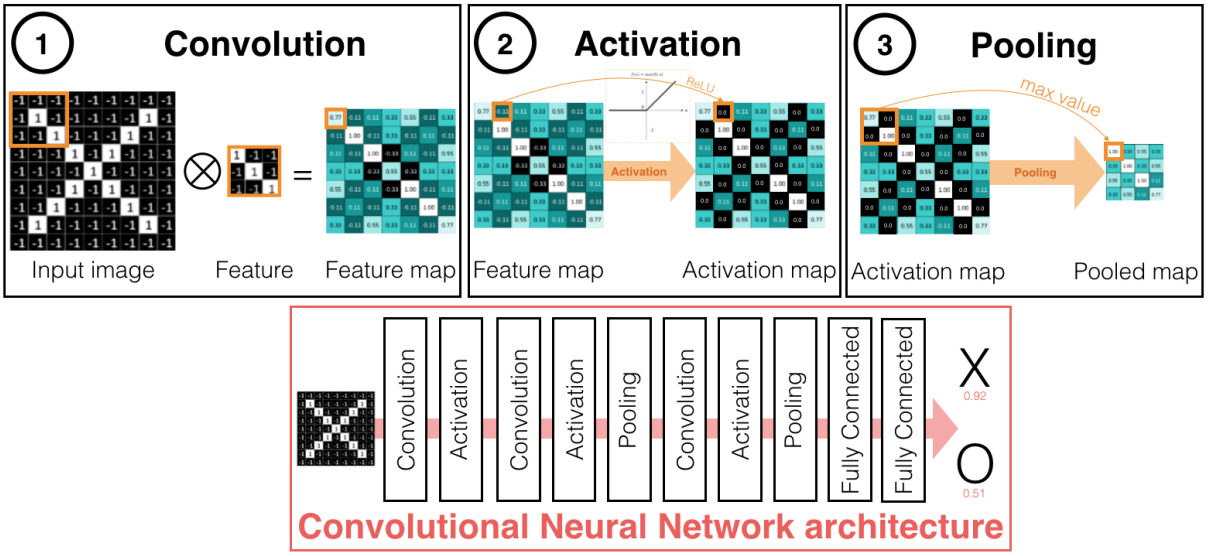


Figure 2.21: Convolutional Neural Network architecture for classification task.

CNN is composed of different layers: Convolution (1), Activation (2), Pooling (3) and Fully Connected layers. (1) The input image is convolved by different filters to produce a feature map. (2) An activation map is applied on each feature map, here the ReLU function is used, setting negative values to zero. (3) Pooling operation is applied on activations maps (e.g. pool maximum value within 2x2 orange kernel). In a fully connected layer, each neuron is connected to every neuron in the previous layer and each connection has an own weight. Figure inspired from https://brohrer.github.io/how_convolutional_neural_networks_work.html

2.5.1.1 Convolutional layer

A convolutional layer can be understood by imagining a flashlight (orange square on **Figure 2.21 1**), sliding across all the input image. The "coverage of the flashlight" (here 3x3) is called "receptive field", whereas the "flashlight" is a filter, which is sliding/convolving. The filters are actually arrays of numbers representing weights. The convolution operation is a simple multiplication between filter values and the pixels within the receptive field. In other words, the convolution consists of a linear combination of the pixels "lighted by the flashlight". In our illustrated example (**Figure 2.21 1**), the filter consists of a diagonal line detector. Resulting values close to 1 mean strong matches, values close to -1 mean strong matches with the "negative version of the input", and values close to zero mean no match at all. After repeating this process for every location on the input image, we obtain a feature map. Similarly, we obtained a stack of feature maps, computed with different filters, resulting in a set of filtered images.

Note that, unlike a fully connected layer, each neuron of a convolutional layer is only connected to a few nearby neurons in the previous layer (see **Figure 2.20**). Besides the advantage of capturing local context, it allows a reduction of memory (i.e. weights) and computation (i.e. connections) needed compared to a fully connected layer.

2.5.1.2 Activation layer

The activation layer encourages an efficient propagation of relevant information, by highlighting signals which are strongly associated with the previous filters. In other words it adjusts how the signal flows from one layer to the next one. This regulation is performed by an activation function, such as Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) (see **Figure 2.21 2**). ReLU is a nonlinear function: setting 0 to all negative values of the feature maps, which helps the CNN for not getting stuck close to 0 or explode toward infinite values.

2.5.1.3 Pooling layer

The pooling layer aims at preserving the most relevant information of the feature maps while reducing the number of parameters (i.e. size of the feature maps). Feature map values of neighbourhoods (e.g. 2x2) are aggregated using a permutation invariant function, such as the max operation (see **Figure 2.21 3**). As a result, the output is insensitive to the exact positive of each feature (translational invariant) but keeps the information whether the features are present in a special neighbourhood. Indeed, only the best fits of each filter are stored within the resulting map.

2.5.1.4 Deep architecture:

The previously presented layers are stacked resulting in a deep architecture (see bottom row of **Figure 2.20**). Input image is filtered, activated, pooled, to create a set of feature-filtered maps, which are filtered, activated, pooled...etc. The first features maps are low level features (e.g. edge detector). But, by convoluting these feature maps, we end up with abstract patterns in the deepest layers, which could have a strong discriminative power.

2.5.1.5 Loss layer

The loss layer is the last layer of a CNN. During the training step, the loss layer shepherds the weights optimization by evaluating the predictions of the CNN on the training samples based on the ground truth. Most of the time, this layer is a ***Fully Connected layer*** (Long et al., 2015). The last fully connected layers translate the deepest feature maps into decision, based on the

weights (i.e. connection strengths) between each value of the maps and each target class (e.g. 0.92 for "X" and 0.51 for "O" in the example of the **Figure 2.21**).

2.5.2 CNNs Training

In this section, we will present some key aspect of the CNNs' training (Bengio, 2012).

2.5.2.1 Backpropagation

At the beginning, the weights and filter values are initialized randomly: the training step aims at optimizing these values based on the training images coupled with their labels, using the backpropagation algorithm (Goodfellow et al., 2016; Rumelhart et al., 1986). This algorithm is in 4 steps and is illustrated in the **Figure 2.22**:

1. *forward step*: a training image pass through the whole network
2. *loss function*: it compares the network output (i.e. predicted label) with its ground truth (i.e. true label). The goal of the backpropagation is to minimize the loss function: we want to find out which weights most directly contributed to the loss of the network
3. *backward pass*: the weights of the network are adjusted accordingly to the obtained loss value, which will be "propagated" through the network, by taking a derivative of the loss with respect to the weights
4. *weight update*: the weights are updated so that they change in the opposite direction of the gradient

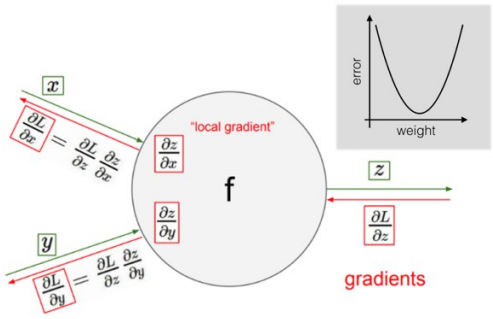


Figure 2.22: Backpropagation algorithm.

The forward step (green arrows) computes z as a function f(x,y), based on the input variables x and y. During the backward step (red arrows), the loss L is back-propagated through the gradient of the loss function with respect to z. The gradients of x and y on the loss function are then computed by applying the chain rule. Figure adapted from CS231n.

There are several methods for performing the parameter update using the gradients, called optimization algorithms, which may be categorized into two classes. The first one manipulates the learning rate globally and maintains it for all weight updates, for the entire training, such as the "Stochastic gradient descent". The second category instead adapts the learning rate depending on the parameters, such as Adam (Kingma and Ba, 2014), which computes adaptive learning rates for different parameters using estimations of first and second moments of the gradients.

2.5.2.2 Learning rate:

The learning rate parameter plays a key role in the backpropagation, and have to be carefully adjusted. As shown by the **Figure 2.23 1**, a high learning rate could prompt a fast convergence (because bigger steps in the weight updates) of the model, but might reach a ceiling (green curve), as if the parameters were bouncing in the optimization space. On the contrary, a low learning rate yields slow convergence (blue curve).

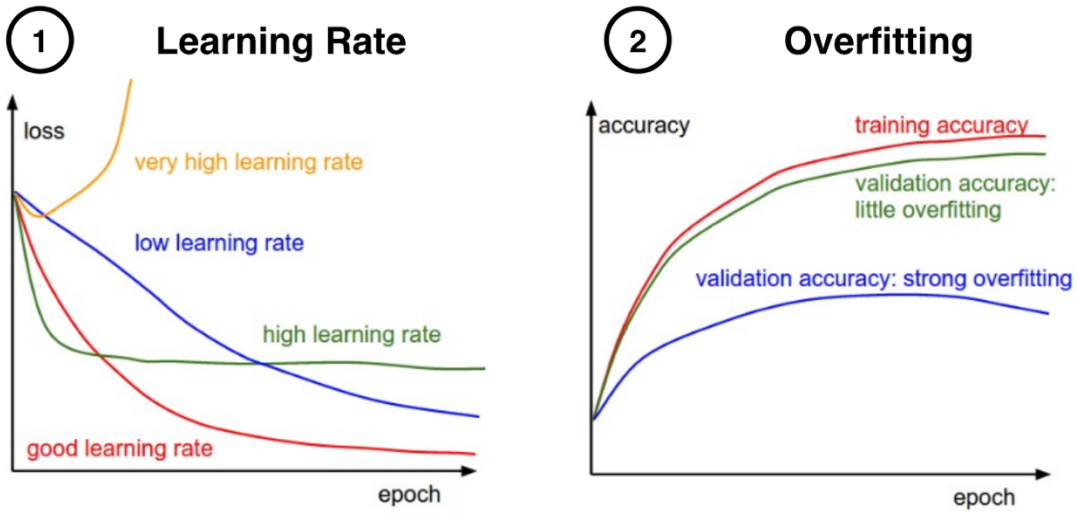


Figure 2.23: Deep learning training.

(1) shows the consequences of the learning rate value on the loss function across the training epochs. (2) illustrates two possible cases of training/validation accuracy across training epochs: little (green) or large overfitting (blue), assessed by the gap between the training and the validation accuracy curves. Note that the epochs (x-axis) count how many times every training sample has been seen by the network. Figure adapted from http://cs231n.github.io/neural-networks-3/.

2.5.2.3 Overfitting

Overfitting occurs when a model learns the details and noise in the training samples, which could negatively impact its ability to generalize to new data. In other words, the model tends to consider random fluctuations in the training data as discriminative features. By too closely fitting to training data, the model limits its predictive power on new data since this "learnt-noise" is not necessarily present in it. During CNN training, overfitting can be observed comparing the training and validation accuracy curves across epochs, as shown in **Figure 2.23 2**. To deal with overfitting, an option is to increase the training dataset (to introduce more variability) or to increase the regularization (see next **subsections 2.5.2.4** and **2.5.2.5**).

2.5.2.4 Dropout

Dropout is a regularization approach commonly used to prevent overfitting when training CNN (Srivastava et al., 2014). Simply said, dropout technique randomly "switches off" some neurons during the training step. In other words, these neurons are not considered during a given forward or backward step. Hence, this method helps reducing inter-dependent learning amongst neurons. Dropout forces the network to learn features that have a discriminative power in combination with many different random subsets of neurons. **Figure 2.24** illustrates this method.

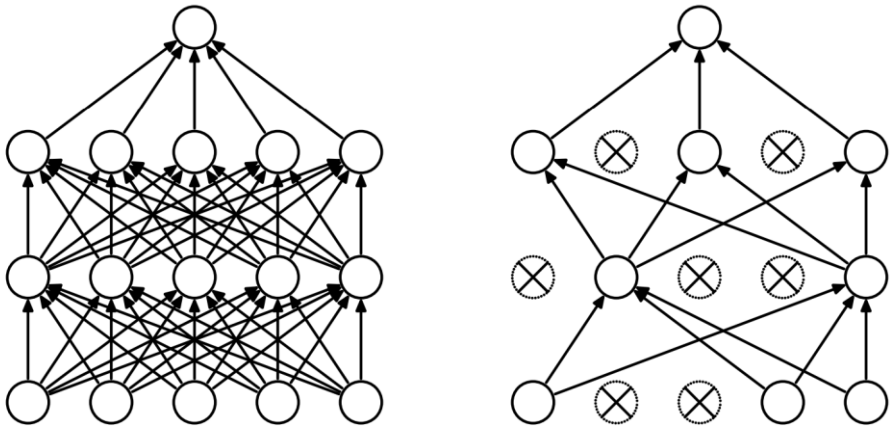


Figure 2.24: DropOut for a Neural Network.

For each training iteration and for each training sample, neurons are either dropped out (i.e. incoming and outgoing connections removed), with a probability 1-p (with p a hyperparameter), either kept, with a probability p. Figure extracted from (Srivastava et al., 2014).

2.5.2.5 Data augmentation

Data augmentation helps to overcome the problem of limited diversity and quantity of data (Goodfellow et al., 2016). Data augmentation consists of inflating the training dataset, by manufacturing new data with the existing available data. Data augmentation techniques apply minor transformations to training data, such as rotations, flips, scaling, noise adding or translations. As a result, data augmentation helps in (i) preventing overfitting by providing more diverse training samples, (ii) making the model more invariant and increasing its generalization ability to new conditions (Dosovitskiy et al., 2014).

In particular, elastic transformations (Simard et al., 2003) have been widely used in medical image segmentation tasks (Oliveira et al., 2017; Perone et al., 2017; Ronneberger et al., 2015; Zaimi et al., 2018). The idea is to synthesize realistic data in order to generate additional data as well as to learn invariance to such plausible tissue deformations.

2.5.3 CNNs for biomedical image segmentation

In this subsection, we will focus on the use of CNNs for the segmentation task while, in the previous paragraphs, we used the classification task for the sake of simplicity. The goal of the segmentation is to assign to each pixel of an image a specific label (e.g. “grey matter”, “white matter”, “lesion”), while the classification task aimed at assigning one label to an entire image (e.g. “healthy brain”, “diseased brain”). A direct application of the previously presented CNN architecture is not relevant for the segmentation task. Indeed, the pooling layers gradually decrease the feature map size: the resulting information contained in the last layer may be too coarse spatially to allow an accurate segmentation. To overcome this issue, several solution were proposed, such as:

- Long et al. proposed a “*Shift-and-stitch*” method: the CNN is applied to shifted versions of the input image and its outputs are stitched together (Long et al., 2015).
- Ronneberger et al. used up-convolutions to increase the image size combined with skip-connections (Ronneberger et al., 2015). It results in a contractive (down-sampling) and a expansive (up-sampling) paths, with skip connections directly connecting them. This architecture will be further detailed in the next subsection.

2.5.3.1 U-net

The most well-known and widely used CNN architecture for medical image segmentation is most probably the U-net, published by Ronneberger et al. in 2015 (Ronneberger et al., 2015). At the time of this writing, the paper ranks the 2nd top-cited paper initially presented during MICCAI conference since its creation, 20 years ago [Source: Nicholas Ayache slides during the the 20th MICCAI conference]. **Figure 2.25** presents its architecture.

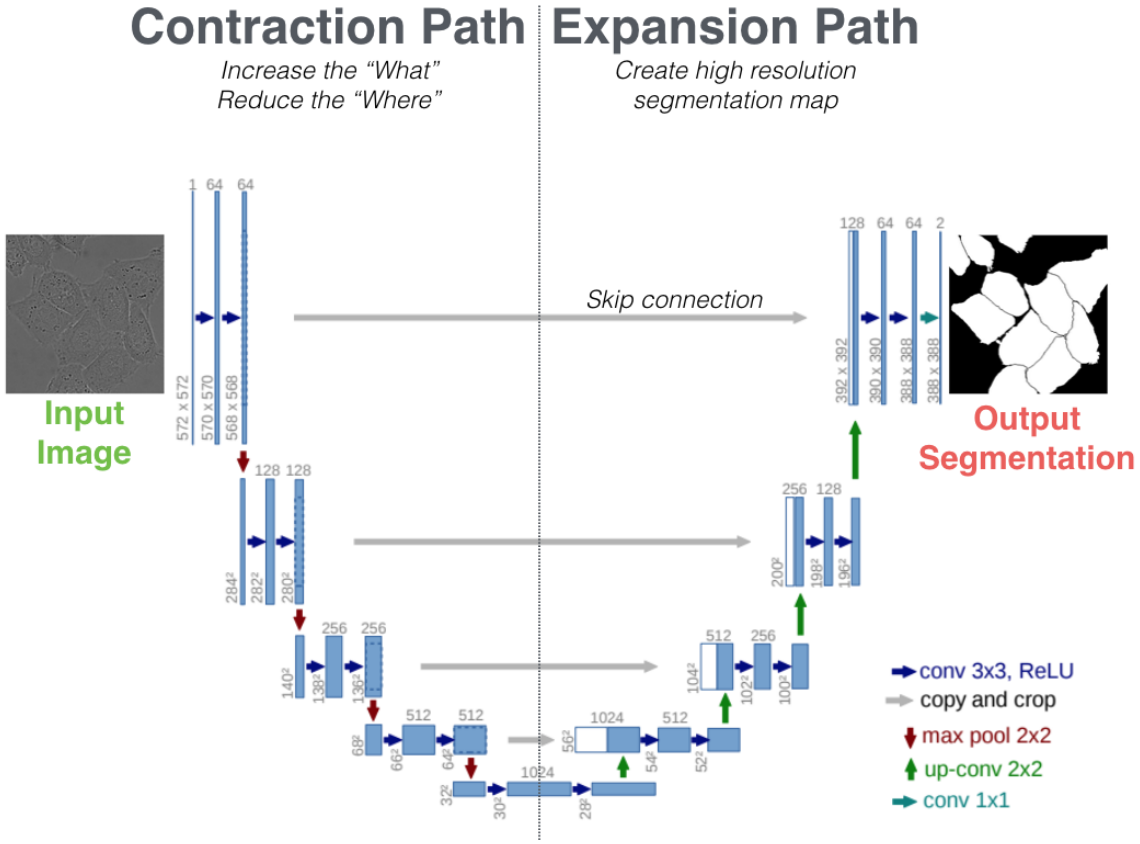


Figure 2.25: U-net architecture.

This network is fed with an input image (left) and output a high-resolution segmentation of it (right). It can be described by a contracting path (left side) where the resolution decreases up to the deepest layer, and a expansion side (right side) which increases the resolution up to the final segmentation. Along the expansion path, skip connections (grey arrows) concatenate features maps from the contracting path (white box) with the feature maps coming from the bottleneck (blue box). Figure adapted from (Ronneberger et al., 2015).

Compared to patch-based methods, the entire image can be directly processed by the U-net, thanks to the combination of downsampling and upsampling layers. This allows the network

to learn from the full spatial information. During the contracting path, the number of feature channels is progressively increased (i.e. increase the "*what*") while the max-pooling reduces the spatial information (i.e. decrease the "where"). During the expansion path, up-convolution operations gradually create a high-resolution segmentation map. Skip connections are used between opposing convolution and deconvolution layers to concatenate features from the contracting and expanding paths, while avoiding the compressive bottleneck of the network. In that way, the network learns compressed latent representations of the data, enriched by contextual representations of the same size.

2.5.3.2 Dilated convolutions

A key downside of the U-net is the important increase of parameters (and then computational cost) induced by the use of deconvolution layers to recover the original image resolution. In addition, the gradual reduction of the receptive fields by the successive max-pooling operations limit the knowledge of the global context. Note that 'receptive field' refers to the area of the input space that impacts a particular neuron of the network.

Yu et Koltun proposed the use of dilated convolutions (also called "à-trous convolutions" or "context module") to exponentially increase the receptive fields, without a huge increase of parameters (Yu and Koltun, 2015). Briefly, dilated convolutions expand the kernel by inserting "spaces" between the kernel elements (i.e. there are d-1 spaces between kernel elements, where d is the dilatation rate). In this way, we increase the receptive field of output units without increasing the kernel size. **Figure 2.26 1** shows how their use can be especially valuable when several dilated convolution layers are stacked one after another. Thus, dilated convolution allows an exponential expansion of the receptive field without loss of coverage, while the number of parameters grows linearly since the number of parameters associated with each layer is identical.

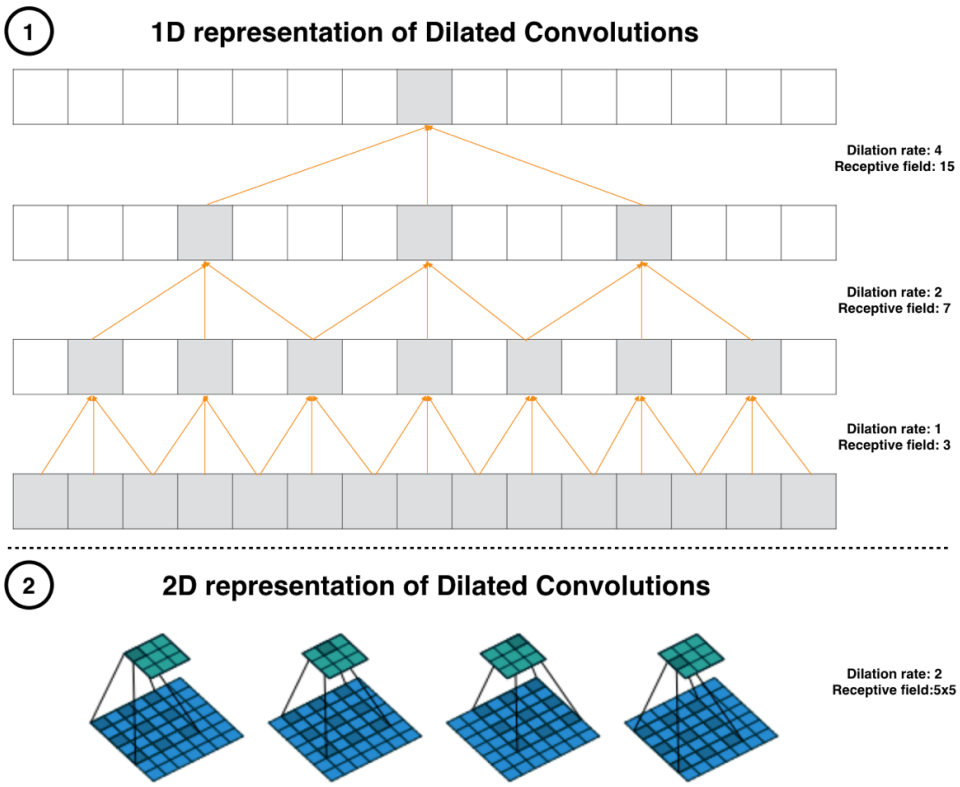


Figure 2.26: Dilated convolutions.

(1) is a 1D representation of dilated convolution stacked one after another. The first layer represents typical convolutions with a kernel size of 3. The above layer represents dilated convolution with a dilation rate of 2, while the top layer convolutions are dilated with a rate of 4. As a result, the effective receptive field of the top convolutions in the bottom layer has a size of 15. (2) shows dilated convolution on 2D data with a dilation rate of 2: we can benefit of a larger receptive field (5x5 vs. 3x3) without loss of coverage.

2.5.4 High class imbalance

A typical issue in real life applications of CNN is that some classes are significantly more represented than other classes. This problem is commonly called class imbalance. In medical image analysis, it is not uncommon that the region of interest occupies only a very small region of the image. It is especially the case with MS lesions since they represent a very small portion of the spinal cord, which is already a small structure. Moreover, a substantial portion of patients with a MS diagnosis has no spinal lesions, which worsen the class balance. Designing deep learning approaches that are adapted to handle this class imbalance is an important research area (Buda et al., 2017; Havaei et al., 2015; Litjens et al., 2017b; Pereira et al., 2016; Sudre et al.,

2017). Indeed, class imbalance could cause the learning process to get stuck in local minima of the loss function, which could lead to predictions drastically biased towards the majority class. As a consequence, the foreground voxels are often missed or only partially detected. Several strategies have been proposed in the past few years to deal with this issue. The main employed strategies were:

2.5.4.1 Oversampling the minority class and/or undersampling the majority class

The straightforward approach to address high class imbalance issue would be to oversample the minority class and/or undersample the majority class. However, it has to be used carefully since (i) oversampling the minority class may lead to overfitting because of the duplicated samples, (ii) undersampling the majority class could reduce the discriminative power of the network since important samples might be missed, (iii) it biases the network towards under-represented classes and may result in over-segmentation (Buda et al., 2017). To overcome this bias, Havaei et al. devised a cascade of two CNN, where the first CNN is trained with balanced classes while the second is trained with proportions near the originals (Havaei et al., 2015). Additionally, they fine-tuned the last output layer on original class balance by re-training only this layer. Thus, they benefited from balanced training samples while ensuring the output probability to be calibrated in accordance with the real class balance.

2.5.4.2 Applying specific data augmentation techniques on positive samples

Instead of doing over and/or under sampling, recent works using CNN-based techniques resorted to specific data augmentation techniques on positive samples (Kamnitsas et al., 2017; Pereira et al., 2016). For example, Pereira et al. applied data augmentation algorithms only on patches containing voxels of the under-represented class. They extensively validated data augmentation techniques (e.g. rotation or scaling transformations) to generate new "realistic brain lesions" in order to deal with class imbalance (Pereira et al., 2016).

2.5.4.3 Using loss functions robust to class imbalance

To alleviate the class imbalance, several techniques proposed to use a specific loss function. Some of them are based on the class balance of the training samples which is used for applying a weighting: background voxels are given less importance than foreground voxels (Brosch et al. 2015; Ronneberger et al. 2015). For instance, Ronneberger et al. (1) adjust the cost function from each class and then (2) correct it by estimating the difficulty of segmenting each

pixel thanks to weight maps which give more importance to some pixels during the training (Ronneberger et al., 2015). However, these functions need to assign weights to samples of different classes to establish the correct balance between foreground and background voxels, which add another hyperparameter to optimize. Sudre et al. compared 4 loss functions designed for unbalanced datasets, and found that loss functions based on overlap measures (e.g. Dice or Generalized Dice Loss) appeared more robust (Sudre et al., 2017), which motivated us to detail the Dice Loss (Milletari et al., 2016) in the next paragraph.

Milletari et al. (Milletari et al., 2016) proposed a loss function inspired by dice coefficient (Dice, 1945). Dice coefficient has value between 0 and 1, where Dice is 0 for disjoint areas, and 1 for perfect agreement, and can be defined as:

$$Dice = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}$$

Where p_i and g_i are the predicted and ground-truth voxels respectively, of the 3D volume to segment (N voxels). By differentiating this formula, we can obtain the associated gradient with respect to the j-th voxel of the prediction:

$$\frac{\partial Dice}{\partial p_j} = 2\frac{g_j(\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2) - 2p_j(\sum_{i=1}^{N} p_i g_i)}{(\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2)^2}$$

Then we can write the Dice loss as:

$$L_{Dice} = -2\frac{\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \varepsilon}$$

With ε a fixed number to ensure the loss stability by avoiding the numerical issues. Note that the minimum value (i.e. the optimal value) is -1.

The main advantages of the Dice loss are: (i) independent of the class balancing, then is adapted to class imbalanced dataset, (ii) it does not require any class frequency balancing as traditional binary cross-entropy losses, (iii) Dice is one of the most used metrics to assess medical image segmentation performance.

CHAPTER 3 METHODOLOGY

This chapter proposes to give some details about the methodology and the progression followed during this Master's project. Herein, the different objectives of my research involvement are enumerated as well as the scientific communications arising from it. **Section 3.1** details the pathway that led to the fully-automatic framework of intramedullary MS lesions in **Chapter 4**. **Section 3.2** briefly presents how the automatic tools were integrated in a major open-source project and within analysis pipelines of clinical data.

3.1 Automatic tools for the segmentation of spinal cord MS lesions

3.1.1 Automatic detection of the spinal cord

To achieve the segmentation of the spinal cord or the MS lesions within it, a first essential step is to detect the spinal cord in the 3D MRI data. Basically, this task aims at detecting the center of the spinal cord and to output its centerline. This centerline is subsequently used for segmentation tasks.

3.1.1.1 Support vector machine and minimal path algorithm

The first implemented and validated solution was using Support Vector Machine (SVM) algorithm, fed with Histogram of Oriented Gradients (HoG) (Dalal and Triggs, 2005), computed on cross-sectional 2D patches. The centerline is then extracted from the resulting heatmap by considering the problem as a "shortest path problem" along the Superior-to-Inferior axis. The problem was solved using a graphical representation, where the nodes are weighted by the SVM values and the edges by the Euclidean distance between the voxels, with a version of Dijkstra's algorithm. This method was validated against a dataset of 195 subjects, and compared with (1) the Hough transform approach used by (De Leener et al., 2014) and (2) a CNN approach, implemented by *Benjamin De Leener*. This work was submitted and accepted for a poster presentation at the *ISMRM conference (2017)*:

- **Gros C.**, Benjamin De Leener, Allan R Martin, Michael G Fehlings, Virginie Callot, Nikola Stikov, Julien Cohen-Adad ; Robust and automatic spinal cord detection on multiple MRI contrasts using machine learning
    - Accepted abstract:

https://drive.google.com/file/d/1GyqxKl6WrAJw5EOBYs4W6RgsJAEBGEdk/view?usp=sharing

3.1.1.2 … using support vector machine and OptiC algorithm

Despite the good encouraging results of the previous approach, further improvements were needed. In particular, the minimal path algorithm was slow and not robust enough to false positive detection of the SVM. *Michaël Sdika*, a collaborator from Lyon (France), implemented an algorithm to overcome the drawbacks of the previous "minimal path algorithm". His implementation (called OptiC) is distance transform based global optimization which features an efficient search (with a linear complexity in the number of voxels) and ensures the global minimum is reached. I adapted it for our specific problem of spinal cord detection to compromise the trade-off between the localization map (computed by a linear SVM trained with HoG features) and the cord continuity along the superior-inferior axis. This approach was validated on a multi-center dataset of 441 subjects and compared with a previously published method, based on Hough transform (De Leener et al., 2014). This work was submitted and accepted for a poster presentation at the *MICCAI conference (2017)*:

- **Gros C.**, Benjamin De Leener, Sara M. Dupont, Allan R. Martin, Michael G. Fehlings, Rohit Bakshi, Subhash Tummala, Vincent Auclair, Donald G. McLaren, Virginie Callot, Michaël Sdika, Julien Cohen-Adad ; OptiC: Robust and Automatic Spinal Cord Localization on a Large Variety of MRI Data Using a Distance Transform Based Global Optimization
    - Accepted paper: https://link.springer.com/chapter/10.1007/978-3-319-66185-8_80
    - Presented poster: https://drive.google.com/file/d/19vyRO1pnzBNVFB-pkZxTQNww8uYUjeEZ/view?usp=sharing

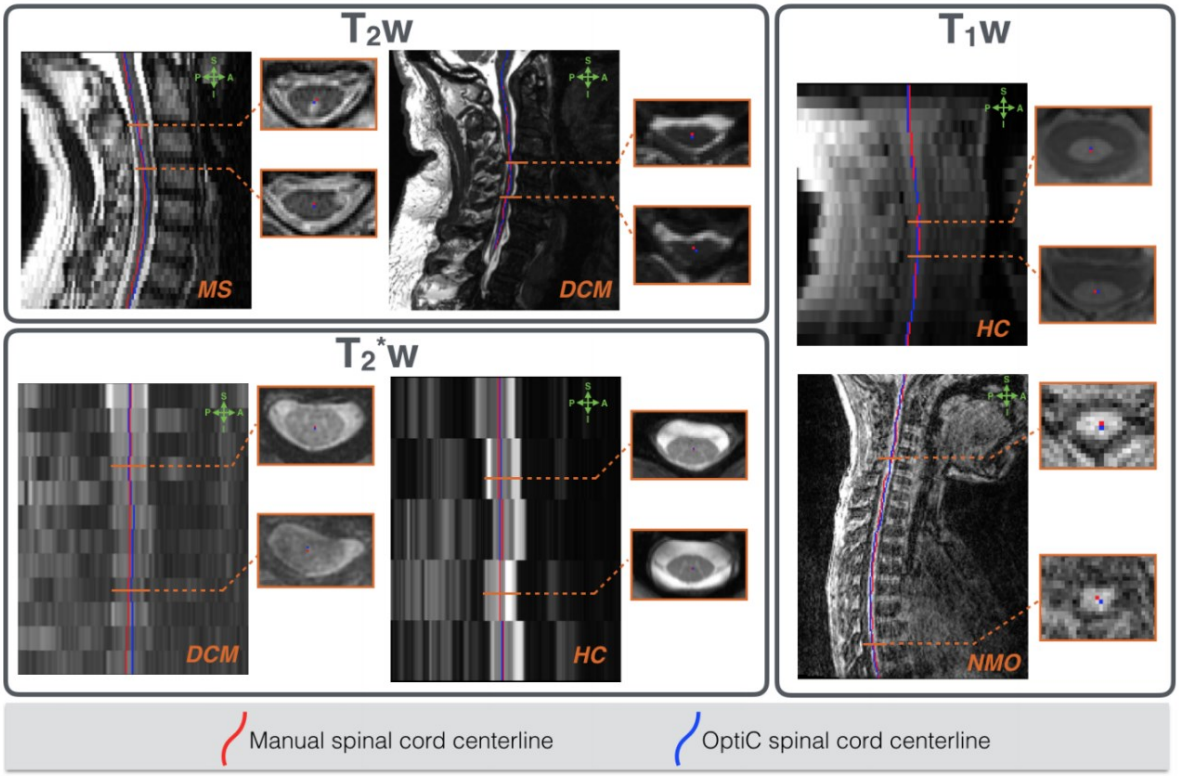


Figure 3.1: Centerline Detection.

Figure extracted from (Gros et al., 2017): "*Examples of automatic SC centerline localization on HC and patients, on T_2-w (top), T_2^*-w (bottom) and T_1-w (left) images. Comparison between ground-truth (red) and automatic (blue) SC centerline on axial and sagittal views.*"

This approach provided very good results on a multi-contrast/pathology/center dataset with mean square error of 1.21 mm when compared to the manually generated centerline. An important point to be noted is that the presented {SVM+HoG} model was trained with only one 3D volume, since no significant improvements were observed when the training involved 5, 10, 15, 20 or 25 volumes. Moreover, the computational time is short: ~30s for a typical T_2-weighted image (including reading and writing tasks, on a standard iMac computer). However, we noted that this approach performed less well on 2D sagittal data (compared to 3D isotropic or 2D axial data).

3.1.1.3 CNN and OptiC algorithm

To further improve the results obtained by the previously presented approach, I then investigated the benefit of initializing OptiC algorithm with a localization map computed by a CNN, instead of a {SVM+HoG} technique. I especially put efforts in improving results on

anisotropic sagittal data. The methods and results of this approach will be detailed in **section 4.2.2** and **4.3.1** respectively.

3.1.2 Automatic segmentation of the spinal cord

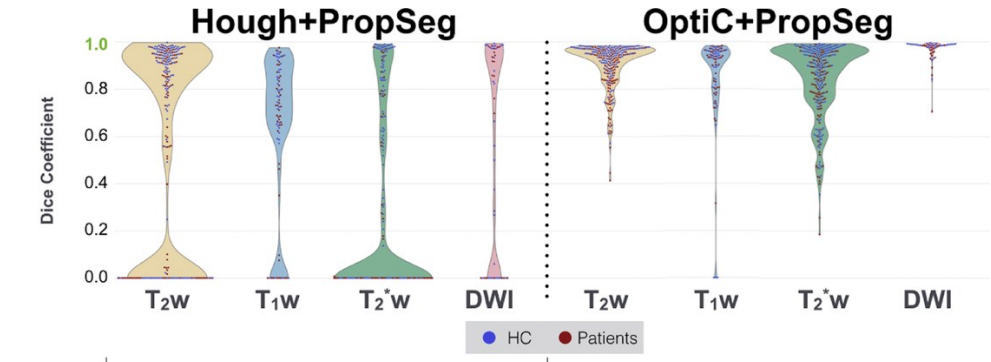
Once the centerline of the spinal cord is detected, the next step towards the segmentation of MS lesions is to automatically define the contours of the spinal cord.

3.1.2.1 PropSeg initialized with {SVM+OptiC} algorithm

PropSeg is an automatic spinal cord segmentation method based on deformable models initialized by Hough transform cord detection (De Leener et al., 2014). As noticed in section 2.2.4.3., we observed that *PropSeg* failed on some instances mainly due to an inaccurate spinal cord detection. Given that {SVM+OptiC} detection approach outperformed Hough transform technique (Gros et al., 2017), we evaluated the improvements of *PropSeg* segmentation when initialized with the centerline produced by {SVM+OptiC} algorithm. We compared the two methods on a multi-center dataset including 501 subjects with a variety of pathological cases: the resulting Dice coefficient (mean ± standard deviation) improved from $44 \pm 43\%$ {*Hough+PropSeg*} to $87 \pm 14\%$ {*OptiC+PropSeg*}, see **Figure 3.2**. This study was reported in an article published in *Medical Image Analysis* (2018) journal:

- **Gros C.**, Benjamin de Leener, Sara M. Dupont, Allan R. Martin, Michael G. Fehlings, Rohit Bakshi, Subhash Tummala, Vincent Auclair, Donald G. McLaren, Virginie Callot, Julien Cohen-Adad, Michaël Sdika ; Automatic spinal cord localization, robust to MRI contrasts using global curve optimization
    - Accepted article:

      https://www.sciencedirect.com/science/article/pii/S136184151730186X



Method	Hough+PropSeg				OptiC+PropSeg			
Contrast	T_2w	T_1w	T_2^*w	DWI	T_2w	T_1w	T_2^*w	DWI
Dice coefficient [avg.] (Best value: 1.0)	0.55 ± 0.43 HC: 0.65±0.42 Patients: 0.41±0.42	0.64 ± 0.32 HC: 0.71±0.28 Patients: 0.59±0.33	0.25 ± 0.38 HC: 0.31±0.41 Patients: 0.15±0.31	0.57 ± 0.43 HC: 0.55±0.42 Patients: 0.61±0.43	**0.90 ± 0.10** HC: 0.93±0.07 Patients: 0.85±0.12	**0.85 ± 0.20** HC: 0.83±0.27 Patients: 0.87±0.12	**0.84 ± 0.16** HC: 0.87±0.14 Patients: 0.78±0.17	**0.96 ± 0.05** HC: 0.98±0.03 Patients: 0.95±0.06

Figure 3.2: OptiC + PropSeg.

Figure extracted from (Gros et al., 2018): "*Results of the spinal cord (SC) segmentation in terms of Dice coefficients between the ground-truth and predicted SC segmentation. For each metric, the optimal value is indicated in green on the y-axis. Results are a comparison of PropSeg (De Leener et al., 2015) segmentations when centerline initialization is performed with the default method (left distribution plots) vs. the OptiC method (right distribution plots), where each point represents a healthy control (blue) or patient (red). Each method was evaluated on the 4 contrast datasets: T_2w, T_1w, T_2*w and DWI (from left to right). The table presents the averaged metrics per contrast for each method.*" The number of total failure of "Hough+PropSeg" is likely due to an inability to detect the center of the spinal cord when using Hough transform.

When using {SVM+OptiC} detection technique, PropSeg' segmentation is substantially improved compared to when using Hough transform approach. However, it remains some instances where the segmentation of PropSeg fails, especially on pathological cases. Further improvements are then needed.

3.1.2.2 JIM software using {SVM+OptiC} algorithm

JIM (http://www.xinapse.com/) is an image analysis software, commonly used by many academic groups for SC analysis using MRI data (Freund et al., 2013; Schlaeger et al., 2014; Yiannakas et al., 2012). A key tool of JIM software is a semi-automatic method for SC segmentation, based on active surface models that deform a cylindrical tubular surface (Horsfield et al., 2010). It requires the user to initialize the segmentation by manually identifying points that

approximate the SC centerline on various slices. I then tried to show the feasibility of making JIM's SC segmentation method fully automatic by providing an automatic SC centerline (Gros et al., 2017) to supplant manual user intervention. Validation on a dataset of 88 volumes (with 3 contrasts, T_2w: n=44, T_1w: n=24, T_2^*w: n=20) show good accuracy of this automatic version of JIM segmentation (Dice=0.87 ± 0.08). Therefore, this module offers the benefits of automatisation (decreased processing time, without user intervention and bias) without losing performance.

3.1.2.3 Sequence of two CNNs

To further improve the results presented in (Gros et al., 2018), I then have designed a framework based on the sequence of two CNNs. The methods and results of this approach will be detailed in **section 4.2.2** and **4.3.1** respectively.

3.1.3 Automatic segmentation of intramedullary lesions

When the contours of the spinal cord are defined within the 3D MRI data, the final step is to discriminate MS lesion voxels from the normal-appearing spinal cord voxels.

A number of approaches were investigated across these 2 years of Master. The progression across the investigated methods was driven by different factors, mainly: (i) my learning process of segmentation methods from basic intensity-threshold-based method to advanced deep learning techniques, (ii) the influence on my work of people with high expertise in segmentation (special thanks to *Tal Arbel* and *Christian Perone*), (iii) the number of available MS patient data (from 3 subjects in day #1, to ~600 in day #-1).

The first class of investigated methods were unsupervised and mainly based on intensity information. For instance, I tested the application of a basic Expectation-Maximisation algorithm with Gaussian Mixture modeling. However, as explained in **paragraph 2.4.4.2**, these approaches suffer from intensity intra (e.g. bias field in the Superior-to-Inferior axis) and inter-subject variability. A lot of efforts were made to reduce this variability with an extensive pre-processing (see **subsection 2.4.1**) (Coupé et al., 2006; Nyúl and Udupa, 1999; Tustison et al., 2010).

Using supervised methods was originally not possible because of the few available MS data and more importantly the lack of labelling. To overcome this limitation, I considered the possibility to generate synthetic lesions on healthy MRI data, as it was done in the brain (Chard et al., 2010;

Gelineau-Morel et al., 2012; Herskovits, n.d.; Melhem, n.d.), and especially the well-known BrainWeb interface (Cocosco et al., 1997). However, the obtained results were not really realistic.

When some manual lesion segmentations were available, I started investigating supervised methods based on feature engineering, such as HoG (Dalal and Triggs, 2005), speeded up robust features (Bay et al., 2008/6), scale-invariant feature transform (Lowe, 2004) or texture (e.g. Gray Level Co-occurrence Matrix, GLCM, features) features. These features were used to feed different machine learning algorithms like SVM, Random Forest, Multi-Layer Perceptron. The results were however not satisfying. One of the reasons for this failure is certainly the class imbalance (i.e. lesion voxels are highly underrepresented compared to normal appearing spinal cord voxels). To limit the class imbalance, I tried to implement a detection module (Maleki et al., 2012) to discriminate cross sectional slices with lesion from slices without lesion, before segmenting the selected slices. However, I was not able to solve overfitting issues.

Lot of investigations in graphical model were then performed. The motivation behind these approaches was to encode spatial context into the model, possibly with multi-scale modelling. In particular, the following methods were explored and partially or completely implemented: (i) hidden Markov random field model and the expectation-maximization algorithm (Zhang et al., 2001), (ii) conditional random fields (Bhole et al., 2011; He et al., 2004; Karimaghaloo et al., 2012). But my home-made implementations did not provide satisfactory results: not-robust initialization, lack of sensibility, poor generalization to new data features, issues in the convergence of graphical model and difficulties in finding hyper-parameters which generalize well to different acquisition centers.

By facing these subsequent failures, I acquired a strong knowledge of the challenges inherent to the automatic segmentation of MS lesions in the spinal cord. It then convinces me to consider deep learning approaches to face the limitations of feature engineering to handle lesion detection while taking advantage of the now (i.e. September 2017) vast amount of available labeled data. However, some challenges remained: (i) class imbalance, (ii) high variability across data coming from different centers. By combining what I have learnt about pre-processing techniques and specificities of spinal cord challenges *with* CNN techniques, I proposed a first framework to

automatically segment MS lesions, which was submitted, then accepted as oral presentation (10min), to the *ISMRM conference* (2018):

- **Charley Gros**, Benjamin De Leener, Atef Badji, Josefina Maranzano, Ren Zhuoquiong, Yaou Liu, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Paola Valsasina, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Sridar Narayanan, Julien Cohen-Adad ; Automatic MS lesion segmentation in the spinal cord using deep learning
    - Accepted abstract:

    https://drive.google.com/file/d/1ISZbtHfCVBRO0jHTim9jJdcICAR3T6jS/view?usp=sharing

Which was further improved and then submitted to *NeuroImage* journal:

- Charley Gros, Benjamin De Leener, Atef Badji, Josefina Maranzano, Dominique Eden, Sara M. Dupont, Jason Talbott, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Leszek Stawiarz, Jan Hillert, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Vincent Auclair, Donald G. McLaren, Allan R. Martin, Michael G. Fehlings, Shahabeddin Vahdat, Ali Khatibi, Julien Doyon, Timothy Shepherd, Erik Charlson, Sridar Narayanan, Julien Cohen-Adad ; Automatic segmentation of spinal cord and intramedullary MS lesions with convolutional neural networks
    - Submitted manuscript presented in the Chapter 4 of this thesis

3.1.3.1 Segmentation of injury in acute blunt spinal trauma

Spinal cord injury (SCI) is damage to the spinal cord which can result from trauma. I collaborated with a research group of San Francisco (USA) to automatically segment SCI lesions

on MRI data, by applying similar techniques than the ones presented in **Chapter 4**. This work resulted in an accepted abstract to *ISMRM conference* (2018):

- McCoy D., Dupont S., **Gros C.**, Julien Cohen-Adad, Jason Talbott ; Convolutional neural network based segmentation of the the spinal cord and intramedullary injury in acute blunt spinal trauma

3.2 Implementation of the automatic tools into analysis pipelines

3.2.1 Integration in the Spinal Cord Toolbox

Spinal Cord Toolbox[1] (SCT) is a comprehensive open-source software dedicated to the processing of spinal cord MRI data (De Leener et al., 2017). It integrates a variety of functions for segmentation (De Leener et al., 2014; Dupont et al., 2017; Perone et al., 2017), template registration (De Leener et al., 2018) and atlas analysis (Lévy et al., 2015). The code is freely available on GitHub[2] and frequently extended with new functions. These tools open the door to robust and reproducible quantitative analysis pipelines of clinical studies (>50 citations).

During my Master, I have implemented some methods related to my research into the SCT: `sct_get_centerline` (automatic SC detection), `sct_deepseg_sc` (automatic SC segmentation), `sct_detect_pmj` (automatic detection of the ponto-medullary junction), `sct_analyze_texture` (compute GLCM features), `sct_analyze_lesion` (compute some lesion characteristics as lesion extension, volume, count in an atlas-based manner).

3.2.2 Grey matter spinal cord atrophy on ALS patients

As detailed in **sub-section 2.3.2**, ALS is a very heterogeneous disease which could benefit from robust biomarkers to better categorize clinical phenotype and improve prognosis. El Mendili et al. established that SC atrophy was a valuable MRI biomarker of disease progression (El Mendili et al., 2014). *Marie-Ève Paquin* led a research project with the following hypothesis: grey matter atrophy is a more sensitive biomarker on ALS patients than the entire SC atrophy. I was involved in this project to (1) devise a pipeline to quantify the grey matter atrophy on ALS patients, (2) evaluate the potential of grey matter atrophy as biomarker to predict ALSFRS score at 1 year. The pipeline used SCT tools (see **Figure 3.2 1.**) and is in open-access (https://osf.io/9xkxx/) to allow its reproducibility by other research groups. We showed that grey matter atrophy was more sensitive to discriminate patients with ALS from healthy controls (see **Figure 3.2 2.**). This study was recently published in the *AJNR* journal:

- Paquin M-E., El Mendili M., **Gros C.**, Dupont SM., Cohen-Adad J., Pradat P.-F. ; Spinal cord gray matter atrophy in amyotrophic lateral sclerosis.

[1] https://sourceforge.net/projects/spinalcordtoolbox/
[2] https://github.com/neuropoly/spinalcordtoolbox

- Accepted article:

  http://www.ajnr.org/content/ajnr/39/1/184.full.pdf

And was disseminated in two symposia:

- **Gros C.**, Paquin M-E. El Mendili M., Dupont SM., Cohen-Adad J., Pradat P.-F. ; Spinal cord gray matter atrophy as MRI biomarker for ALS patients ; *International symposium on ALS/MND-2017* ; Accepted Abstract, Oral presentation (15min)
- Montreal Neuro Symposium: Accepted Poster:

https://drive.google.com/file/d/1KX56Qi6CNifxyFdVbZ7d8rQ4GxDGiVL-/view?usp=sharing

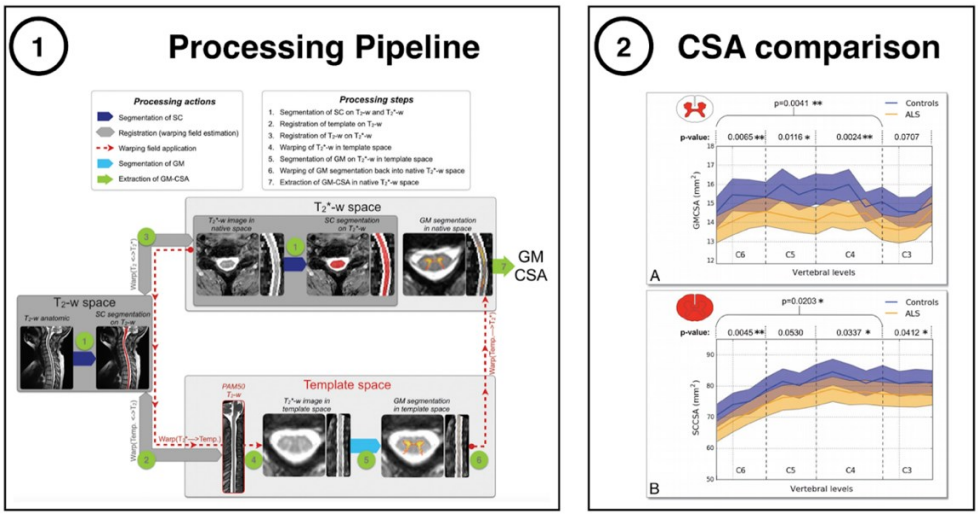


Figure 3.3: Spinal cord grey matter atrophy in ALS patients.

(1) shows the processing pipeline used to measure the grey matter cross-sectional area (GMCSA) based on the cord and white matter segmentation. (2) represents the GMCSA (A) and SCCSA (B) measured on ALS patient (orange) and healthy controls (blue), averaged within group and plot along the cervical SC axis. We observed a stronger intergroup difference for GMCSA, compared to SCCSA. Figures extracted from (Paquin et al., 2018).

3.2.3 Spatial distribution of cervical MS lesions

The preferential spatial distribution of MS lesions within the spinal cord may be considered as a valuable prior information when integrated into an automatic segmentation pipeline, as it was shown to help for brain lesion segmentation (Harmouche et al., 2015). Besides of this and more

importantly, the characterisation of the lesion topography could help to better understand the association of MS involvement in the spinal cord with disease status and progression. *Dominique Eden* fashioned a quantitative study to investigate the spatial distribution of MS lesions and its associations with clinical status. I was involved in this project through the design of an analysis pipeline, intended to loop across about 500 patients from 11 acquisitions centers, in order to (1) generate a Lesion Probability Map, (2) perform an atlas based analysis of the lesion involvement, see **Figure 3.4**. Both code and Lesion Probability Maps are in open-access: https://osf.io/cx5ur/. We notably found that the lesion occupancy was greater in the C1-C3 vertebral levels than the C4-C7 portion, while, among the white matter columns, DC areas were more occupied by lesions than LF and VF areas. We submitted this study to *Brain* journal:

- Dominique Eden and **Charley Gros**, Atef Badji, Sara M. Dupont, Benjamin De Leener, Josefina Maranzano, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Leszek Stawiarz, Jan Hillert, Jason Talbott, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth A. Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Timothy Shepherd, Erik Charlson, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Sridar Narayanan, Julien Cohen-Adad ; Spatial distribution of multiple sclerosis lesions in the cervical cord

And communicated the results in different conferences:

- **Charley Gros** and Dominique Eden, Atef Badji, Sara M. Dupont, Benjamin De Leener, Josefina Maranzano, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Leszek Stawiarz, Jan Hillert, Jason Talbott, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth A. Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Timothy Shepherd, Erik Charlson, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Sridar Narayanan, Julien Cohen-Adad ; Spatial distribution of

multiple sclerosis lesions in the cervical cord ; *ARSEP-2017* ; Accepted Abstract, Oral Presentation (10min).

https://drive.google.com/file/d/1Yynfst_zSjJIX3hVcOgJDpuzXpWWgvFG/view?usp=sharing

- **Charley Gros** and Dominique Eden, Atef Badji, Sara M. Dupont, Benjamin De Leener, Josefina Maranzano, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Leszek Stawiarz, Jan Hillert, Jason Talbott, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth A. Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Timothy Shepherd, Erik Charlson, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Sridar Narayanan, Julien Cohen-Adad ; Spatial distribution of multiple sclerosis lesions in the cervical cord ; *QBIN-2018* (Québec Bio-Imaging Network) ; Accepted Abstract, Poster Presentation.

  https://drive.google.com/file/d/12QuVN36mGwOOHE6U5XqURzcZ_0fPD0lU/view?usp=sharing

- Dominique Eden, **Charley Gros**, Atef Badji, Sara M. Dupont, Benjamin De Leener, Josefina Maranzano, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Leszek Stawiarz, Jan Hillert, Jason Talbott, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth A. Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Timothy Shepherd, Erik Charlson, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Sridar Narayanan, Julien Cohen-Adad ; Spatial distribution of multiple sclerosis lesions in the cervical cord ; *ISMRM-2018* ; Accepted Abstract, Oral presentation (10min).

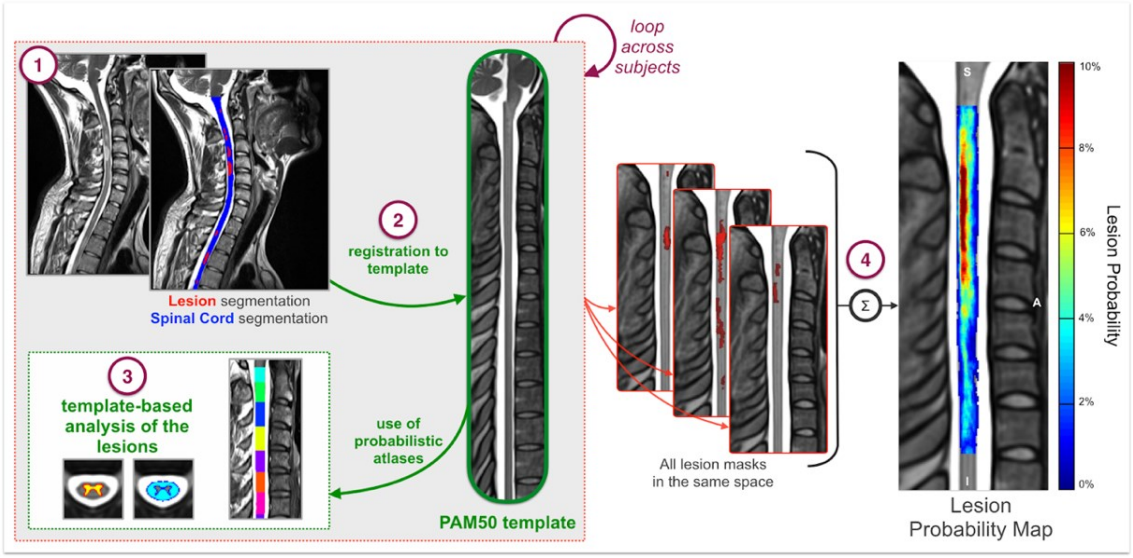


Figure 3.4: Spatial distribution of MS lesions in the cervical spinal cord.

Quantitative analysis pipeline used in [REF][3] with the following steps: (1) generation of cord and binary lesion masks, (2) registration to the PAM50 template, (3) use of probabilistic atlases to compute lesion characteristics, and (4) mapping of weighted lesion masks in the template space and generation of a lesion probability map.

[3] The reference will be updated in a revised version of the thesis as the paper has been submitted but not published yet.

CHAPTER 4 ARTICLE 1: AUTOMATIC SEGMENTATION OF THE SPINAL CORD AND INTRAMEDULLARY MULTIPLE SCLEROSIS LESIONS WITH CONVOLUTIONAL NEURAL NETWORKS

Submitted to NeuroImage, May 8th 2018.

Title: Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks

Authors:

Charley Gros [1], Benjamin De Leener [1], Atef Badji [1,2], Josefina Maranzano [4], Dominique Eden [1], Sara M. Dupont [1,3], Jason Talbott [3], Ren Zhuoquiong [5], Yaou Liu [5,6], Tobias Granberg [7,8], Russell Ouellette [7,8], Yasuhiko Tachibana [22], Masaaki Hori [23], Kouhei Kamiya [23], Lydia Chougar [23,24], Jan Hillert [7], Leszek Stawiarz [7], Elise Bannier [9,10], Anne Kerbrat [10,11], Gilles Edan [10,11], Pierre Labauge [12], Virginie Callot [13,14], Jean Pelletier [14,15], Bertrand Audoin [14,15], Henitsoa Rasoanandrianina [13,14], Jean-Christophe Brisset [16], Paola Valsasina [17], Maria A. Rocca [17], Massimo Filippi [17], Rohit Bakshi [18], Shahamat Tauhid [18], Ferran Prados [19,26], Marios Yiannakas [19], Hugh Kearney [19], Olga Ciccarelli [19], Seth Smith [20], Constantina Andrada Treaba [8], Caterina Mainero [8], Jennifer Lefeuvre [21], Daniel S. Reich [21], Govind Nair [21], Vincent Auclair [27], Donald G. McLaren [27], Allan R. Martin [28], Michael G. Fehlings [28], Shahabeddin Vahdat [29,25], Ali Khatibi [4,25], Julien Doyon [4,25], Timothy Shepherd [30], Erik Charlson [30], Sridar Narayanan [4], Julien Cohen-Adad [1,25]

Affiliations:

[1] NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada

[2] Department of Neuroscience, Faculty of Medicine, University of Montreal, Montreal, QC, Canada

[3] Department of Radiology and Biomedical Imaging, Zuckerberg San Francisco General Hospital, University of California, San Francisco, CA, USA

[4] McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada

[5] Department of Radiology, Xuanwu Hospital, Capital Medical University, Beijing 100053, P. R. China

[6] Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, P. R. China

[7] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[8] Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, USA

[9] CHU Rennes, Radiology Department

[10] Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Visages U1128, France

[11] CHU Rennes, Neurology Department

[12] MS Unit. DPT of Neurology. University Hospital of Montpellier

[13] Aix Marseille Univ, CNRS, CRMBM, Marseille, France

[14] APHM, CHU Timone, CEMEREM, Marseille, France

[15] APHM, Department of Neurology, CHU Timone, APHM, Marseille

[16] Observatoire Français de la Sclérose en Plaques (OFSEP) ; Univ Lyon, Université Claude Bernard Lyon 1 ; Hospices Civils de Lyon ; CREATIS-LRMN, UMR 5220 CNRS & U 1044 INSERM ; Lyon, France

[17] Neuroimaging Research Unit, INSPE, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy

[18] Brigham and Women's Hospital, Harvard Medical School, Boston, USA

[19] Queen Square MS Centre, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, London (UK)

[20] Vanderbilt University, Tennessee, USA

[21] National Institute of Neurological Disorders and Stroke, National Institutes of Health, Maryland, USA

[22] National Institute of Radiological Sciences, Chiba, Chiba, Japan

[23] Juntendo University Hospital, Tokyo, Japan

[24] Hospital Cochin, Paris, France

[25] Functional Neuroimaging Unit, CRIUGM, Université de Montréal, Montreal, QC, Canada

[26] Center for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

[27] Biospective Inc., Montreal, QC, Canada

[28] Division of Neurosurgery, Department of Surgery, University of Toronto, Toronto, ON, Canada

[29] Neurology Department, Stanford University, US

[30] NYU Langone Medical Center, New York, USA

Acknowledgements:

The following people are acknowledged for MRI acquisition: Manuel Taso, Jamie Near, Ives Levesque, Guillaume Gilbert, Robert Barry, Johanna Vannesjo, Antonys Melek and Charles Tremblay.

The following people are acknowledged for sharing data: Eric Klawiter (Massachusetts General Hospital), Julius Dewald, Haleh Karbasforoushan (Northwestern University), Pierre-François Pradat and Habib Benali (Pitié-Salpêtrière Hospital), Barry Bedell (Biospective), Claudia Gandini Wheeler-Kingshott (University College London), Pierre Rainville (Université de Montréal), Bailey Lyttle, Benjamin Conrad, Bennett Landman (Vanderbilt University), Maryam Seif and Patrick Freund (Spinal Cord Injury Center Balgrist, University Hospital Zurich), Seok Woo Kim, Jisun Song, Tom Lillicrap and Emil Ljungberg.

We acknowledge the NVIDIA Corporation for the donation of a GPU.

We would like to warmly thank the members of NeuroPoly Lab for fruitful discussions and valuable suggestions, especially Harris Nami and Ryan Topfer for reviewing the manuscript and Christian Perone and Francisco Perdigón Romero for their inputs on deep learning.

Grant Support:

Funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging (JCA), the Canadian Institute of Health Research [CIHR FDN-143263], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [28826], the Fonds de Recherche du Québec - Nature et Technologies [2015-PR-182754], the Natural Sciences and Engineering Research Council of Canada [435897-2013], IVADO, TransMedTech and the Quebec BioImaging Network, ISRT, Wings for Life (INSPIRED project), the SensoriMotor

Rehabilitation Research Team (SMRRT), the National Multiple Sclerosis Society NMSS RG-1501-02840 (SAS), NIH/NINDS R21 NS087465-01 (SAS), NIH/NEI R01 EY023240 (SAS), DoD W81XWH-13-0073 (SAS), the Intramural Research Program of NIH/NINDS (JL, DSR, GN), the Centre National de la Recherche Scientifique (CNRS), The French Hospital Programme of Clinical Research (PHRC) for the EMISEP project, ClinicalTrials.gov Identifier: NCT02117375, the “Fondation A*midex-Investissements d'Avenir” and the “Fondation Aix-Marseille Université”, the Stockholm County Council (ALF grant 20150166), a postdoc fellowship from the Swedish Society for Medical Research (TG), a postdoc non-clinical fellowship from Guarantors of Brain (FP), the French State and handled by the "Agence Nationale de la Recherche", within the framework of the "Investments for the Future" programme, under the reference ANR-10-COHO-002 Observatoire Français de la Sclérose en plaques (OFSEP), with the assistance of Eugène Devic EDMUS Foundation against multiple sclerosis; EDMUS, a European database for multiple sclerosis. Confavreux C, Compston DAS, Hommes OR, McDonald WI, Thompson AJ. J Neurol Neurosurg Psychiatry 1992; 55: 671-676, NIH/NINDS R21 NS087465-01 (SAS), NIH/NEI R01 EY023240 (SAS), DoD W81XWH-13-0073 (SAS), Grant MOP-13034, National Multiple Sclerosis Society NMSS RG-1501-02840 (SAS). Additional funding sources include NIH/NINDS R21 NS087465-01 (SAS), NIH/NEI R01 EY023240 (SAS) and DoD W81XWH-13-0073 (SAS).

Corresponding author:

Julien Cohen-Adad
Dept. Genie Electrique, L5610
Ecole Polytechnique
2900 Edouard-Montpetit Bld
Montreal, QC, H3T 1J4, Canada
Phone: 514 340 5121 (office: 2264); Skype: jcohenadad; e-mail: jcohen@polymtl.ca

Abbreviations:

ALS: amyotrophic lateral sclerosis ; **CNN**: convolutional neural network ; **DCM**: degenerative cervical myelopathy ; **HC**: healthy control ; **IQR**: interquartile range ; **MS**: multiple sclerosis ; **MSE**: mean square error ; **NMO**: neuromyelitis optica ; **SCI**: spinal cord injury ; SCT: spinal cord toolbox ; **SVM**: support vector machine ; **SYR**: syringomyelia ; **RVD**: relative volume difference

ABSTRACT

Context: The spinal cord is frequently affected by atrophy and/or lesions in multiple sclerosis (MS) patients. Segmentation of the spinal cord and lesions from MRI data provides measures of atrophy and lesion burden, which are key criteria for the diagnosis, prognosis and longitudinal monitoring in MS. Automating this operation eliminates inter-rater variability and increases the efficiency of large-throughput analysis pipelines. Achieving robust and reliable segmentation across multi-center spinal cord data is challenging because of the large variability related to acquisition parameters and image artifacts. In particular, a precise delineation of lesions is hindered by a broad heterogeneity of lesion contrast, size, location, and shape.

Goal: To develop a fully-automatic framework — robust to variability in both image parameters and clinical condition — for segmentation of the spinal cord and intramedullary MS lesions from conventional MRI data of MS and non-MS cases.

Data: 1042 adult subjects (459 healthy controls, 471 MS patients, and 112 with other spinal pathologies) were included in this multi-center study (n=30 centers). Data spanned 3 contrasts (T_1-, T_2- and T_2^*-weighted) for a total of 1943 volumes and featured large heterogeneity in terms of resolution, orientation, coverage and clinical conditions.

Method: The proposed cord and lesion automatic segmentation approach is based on a sequence of two Convolutional Neural Networks (CNN). To deal with the very small proportion of spinal cord and/or lesion voxels compared to the rest of the volume, a first CNN with 2D dilated convolutions detects the spinal cord centerline followed by a second 3D CNN that segments the spinal cord and lesions. CNNs were trained independently with the Dice loss.

Results: When compared to a state-of-the-art spinal cord segmentation method (PropSeg), our CNN-based approach showed a median Dice of 95 % vs. 88 % for PropSeg. Regarding lesion segmentation, our framework, when compared with manual segmentation of MS patients, provided a lesion-wise detection sensitivity of 83 %, a precision of 77 %, a relative volume difference of 15 %, and a Dice of 60 %.

Conclusion: We introduced a robust method to segment the spinal cord and intramedullary MS lesions on a variety of MRI contrasts. The proposed framework is open-source and readily available in the Spinal Cord Toolbox.

Keywords: MRI, Segmentation, Spinal cord, Multiple sclerosis, Convolutional neural networks

4.1 Introduction

Multiple sclerosis (MS) is a chronic immune mediated disease of the central nervous system, with variable clinical expression. The pathologic hallmark of MS is the occurrence of focal areas of inflammatory demyelination within the brain and spinal cord, known as lesions (Popescu and Lucchinetti, 2012). MS lesions exhibit variable degrees of demyelination, axonal injury and loss, remyelination and gliosis. Impaired axonal conduction often causes motor, sensorial, visual, and cognitive impairment (Compston and Coles, 2002). Clinicians and researchers extensively use conventional MRI (e.g., T2-weighted) to non-invasively quantify the lesion burden in time and space (Filippi and Rocca, 2007; Kearney et al., 2015b; Simon et al., 2006; Sombekke et al., 2013; Weier et al., 2012). The study of spinal cord lesions has recently garnered interest (Hua et al., 2015; Kearney et al., 2015a) given its potential value for diagnosis and prognosis of MS (Arrambide et al., 2018; Sombekke et al., 2013; Thorpe et al., 1996). Moreover, spinal cord atrophy is common in MS (Bakshi et al., 2005) and the quantification of such atrophy is clinically relevant and correlates with clinical disability (Cohen et al., 2012; Kearney et al., 2014; Losseff et al., 1996; Lundell et al., 2017; Rocca et al., 2013, 2011). Consequently, segmentation of the spinal cord and MS lesions contained within it (intramedullary lesions) is a common procedure to quantitatively assess the structural integrity of this portion of the central nervous system in MS patients. However, manual segmentation is time-consuming and suffers from intra- and inter-rater variability. Hence, there is a need for robust and automatic segmentation tools for the spinal cord and the MS lesions within in.

Various automatic spinal cord segmentation methods have been proposed in the past few years, including active contours and surface-based approaches (De Leener et al., 2015; Koh et al., 2010), and atlas-based methods (Carbonell-Caballero et al., 2006; Chen et al., 2013; Pezold et al., 2015; Tang et al., 2013). While these methods have shown good performance (De Leener et al., 2016), they often require a specific region of interest and/or are limited to a specific contrast and resolution. Moreover, the lack of validation against multi-center data or cases with spinal cord damage has limited their application in large clinical multi-site studies. Automatic spinal cord segmentation is difficult to achieve robustly and accurately across the broad range of spinal cord shapes, lengths and pathologies; and across variable image dimensions, resolutions, orientations, contrasts, and artifacts (e.g. susceptibility, motion, chemical shift, ghosting, blurring, Gibbs).

Figure 4.1 illustrates these challenges, depicting the heterogeneity frequently observed in multi-center, clinical, spinal cord datasets.

The automatic segmentation of MS lesions has been thoroughly investigated over the past two decades for brain datasets (García-Lorenzo et al., 2013; Lladó et al., 2012), although it still remains a challenging task (Kamnitsas et al., 2017; Meier et al., 2018; Valverde et al., 2017a, 2017b). While previous methods have shown reasonable performance in the brain, they are not easily transposable to the spinal cord, mainly because of its specific morphology. Furthermore, traditional intensity-based segmentation methods are challenging in spinal cord images because of (i) the frequent intensity bias field in the Superior-to-Inferior axis which is difficult to correct, (ii) the confounding of lesion intensities with those of normal structures (e.g. grey matter on T_2*-weighted images), or artifacts, and (iii) partial volume effects, where several structures may contribute to the signal of border voxels (e.g. cerebrospinal fluid and cord). To provide an overview of these challenges, **Figure 4.1** shows instances of spinal cord MS lesions exhibiting heterogeneity (i.e. location, size and shape), along with their intensity histograms, which demonstrate a large overlap with the spinal cord intensities.

The last years have witnessed a noteworthy interest in convolutional neural networks (CNNs) for image segmentation tasks, with remarkable performance in different domains, notably in medical image analysis (Litjens et al., 2017). The game-changing advantage of CNNs, compared to feature engineering based approaches, is their hierarchical representation learning strategy to find appropriate filters on their own. Indeed, the features learned in the first layers come together and make abstract shapes, which often have meaning in their deep layers. CNN methods have proven to be highly robust to varying image appearances. In particular, since 2015, U-net architecture achieved a notable breakthrough in the biomedical image segmentation community (Ronneberger et al., 2015), even for tasks with little available annotated training data. The good performance of U-net architecture is often explained by the use of two distinct paths: a contracting path to capture context, followed by a symmetric expanding path to recover the spatial information, with the support of skip connections between the paths. However, training CNNs on very unbalanced datasets, such as those encountered in MS spinal cord lesion segmentation tasks (i.e. data with < 1% of lesion voxels), remains a focus of active research (Buda et al., 2017; Christ et al., 2017; Sudre et al., 2017).

In this work, we propose an original and fully automatic framework for segmenting the spinal cord and intramedullary MS lesions from a variety of MRI contrasts and resolutions. The presented methods are based on a sequence of CNNs, trained from scratch and specifically designed for spinal cord morphometry. We validated its robustness against a multi-center clinical dataset ($n_{volumes}$=1943), which features a variety of pathologies, artifacts, contrasts, resolutions, dimensions, and orientations.

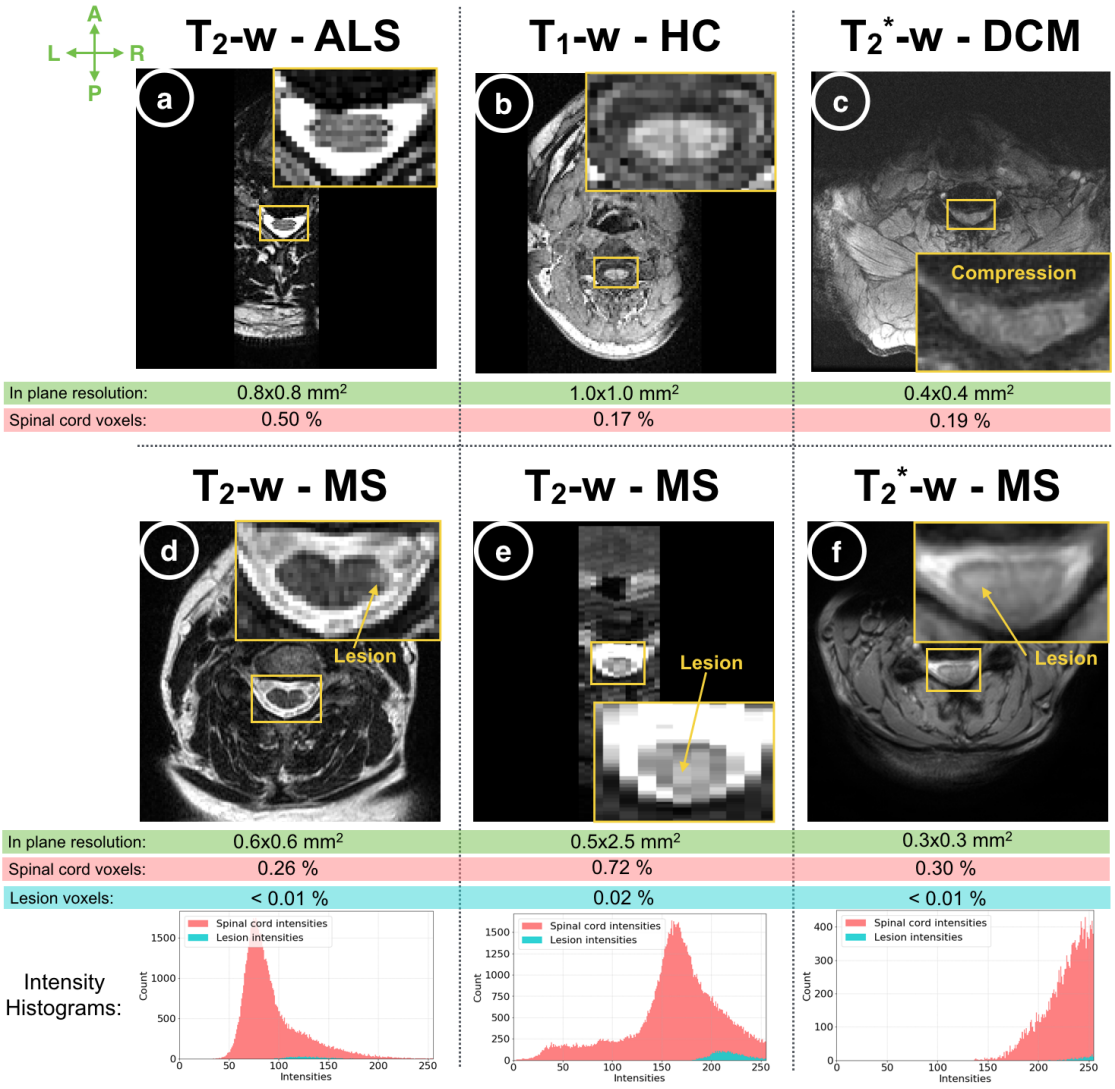


Figure 4.1: Spinal cord axial slice samples.

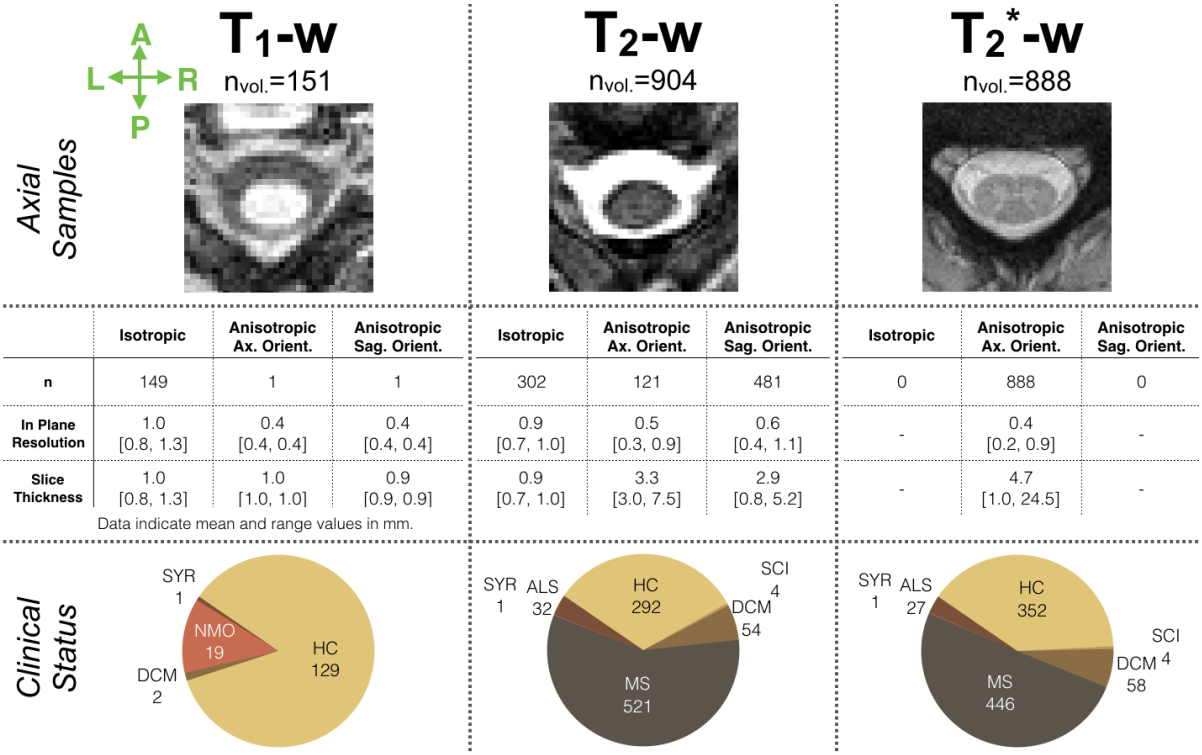
Samples (a-f) show the variability of the images in terms of resolution, field of view and MR contrasts. These images were acquired in 6 different centers, from a variety of subjects in terms of clinical status: healthy control (HC, b), amyotrophic lateral sclerosis (ALS, a), degenerative cervical myelopathy (DCM, c) and multiple sclerosis (MS, d-f). The in-plane resolutions vary across the images. For all images, the spinal cord and lesion voxels represent less than 1 % and 0.1 % respectively of the entire volume. The shape, location, size and level of contrast differ among MS lesions (d-f). The histograms for spinal cord and lesion voxels of the MS patient (d-f) images are shown at the bottom. A substantial overlap between spinal cord and lesion intensities is observed, leading to a low contrast, especially for T_2^*-w images (f) with similarities between grey matter and lesion appearance.

4.2 Material and Method

4.2.1 Data

30 centers contributed to this study, gathering retrospective 'real world' data from 1042 subjects, including healthy controls (n=459), patients with MS or suspected MS (n=471), as well as degenerative cervical myelopathy (n=55), neuromyelitis optica (n=19), spinal cord injury (n=4), amyotrophic lateral sclerosis (n=32) and syringomyelia (n=2). The MS cohort spans a large heterogeneity of clinical conditions in terms of the Expanded Disability Status Scale (mean: 2.5 ; range: 0-8.5) and phenotype: clinically isolated syndrome (n=29), relapsing-remitting MS (n=283), secondary progressive MS (n=76), and primary progressive MS (n=69). Clinical data were not available for all MS patients. Images were acquired at 3T and 7T on various platforms (Siemens, Philips and GE). Contrasts included T_2−weighted ($n_{vol.}$ = 904), T_1−weighted ($n_{vol.}$ = 151) and T_2*-weighted ($n_{vol.}$ = 888). The coverage substantially differed among subjects, with volumes including the brain and/or diverse vertebral levels (cervical, thoracic, lumbar). Spatial resolutions included isotropic ($n_{vol.}$ = 451, from 0.7 to 1.3 mm) and anisotropic data with axial ($n_{vol.}$ = 1010, in plane: from 0.2 to 0.9 mm, slice thickness including slice gap: from 1.0 to 24.5 mm), or with sagittal orientation ($n_{vol.}$ = 482, in plane: from 0.4 to 1.1 mm, slice thickness: from 0.8 to 5.2 mm). **Figure 4.2** summarises the dataset.

4 trained raters (BDL, SD, DE, CG) manually corrected the segmentation produced by *PropSeg* (De Leener et al., 2014) using FSLview (Jenkinson et al., 2012). The resulting spinal cord mask was considered as ground-truth and is herein referred to as "manual segmentation". Using data from MS patients ($n_{vol.}$=967), lesion masks were generated by 7 raters including radiologists (JM, JT, MH, YT, RZ, LC) and trained raters (AB) using ITK-SNAP Toolbox 3.6.0 (Yushkevich and Gerig, 2017). Image raters were blind to diagnostic and clinical information. When no lesions were detected ($n_{vol.}$=171), an empty lesion mask was generated. Guidelines followed by raters are available at: osf.io/d4evy/. The lesion involvement was highly heterogeneous across patients, with a mean (range) lesion count of 3.1 (0-17) and total lesion volume of 192 mm^3 (0.0-1679.8mm^3). Over the entire MS dataset, 0.01 % of image voxels on average were confirmed to contain lesions, showing the unbalanced nature of the data.



	Isotropic	Anisotropic Ax. Orient.	Anisotropic Sag. Orient.	Isotropic	Anisotropic Ax. Orient.	Anisotropic Sag. Orient.	Isotropic	Anisotropic Ax. Orient.	Anisotropic Sag. Orient.
n	149	1	1	302	121	481	0	888	0
In Plane Resolution	1.0 [0.8, 1.3]	0.4 [0.4, 0.4]	0.4 [0.4, 0.4]	0.9 [0.7, 1.0]	0.5 [0.3, 0.9]	0.6 [0.4, 1.1]	-	0.4 [0.2, 0.9]	-
Slice Thickness	1.0 [0.8, 1.3]	1.0 [1.0, 1.0]	0.9 [0.9, 0.9]	0.9 [0.7, 1.0]	3.3 [3.0, 7.5]	2.9 [0.8, 5.2]	-	4.7 [1.0, 24.5]	-

Data indicate mean and range values in mm.

Figure 4.2: Overview of the dataset.

Samples of cross-sectional axial slices of the three MR contrast datasets (T_1-weighted, T_2-weighted, T_2^*-weighted) are depicted (top row). Image characteristics in terms of orientation (orient.) and resolution (resol.), grouped by isotropic, anisotropic and with axial (Ax.) orientation or sagittal (Sag.) orientation are presented (middle row). The last row shows the proportion of clinical status among the imaged subjects, including: healthy controls (HC), multiple sclerosis (MS), degenerative cervical myelopathy (DCM), neuromyelitis optica (NMO), traumatic spinal cord injury (SCI), amyotrophic lateral sclerosis (ALS), and syringomyelia (SYR).

4.2.2 Segmentation framework

The proposed segmentation framework is depicted in **Figure 4.3**. The workflow consists of two major stages. The first stage aims at detecting the spinal cord, predicting a spinal cord centerline (steps 1 and 2 in **Figure 4.3**). The second stage performs the spinal cord and lesion automatic segmentation, independently, along the previously predicted centerline (steps 3 and 4 in **Figure 4.3**).

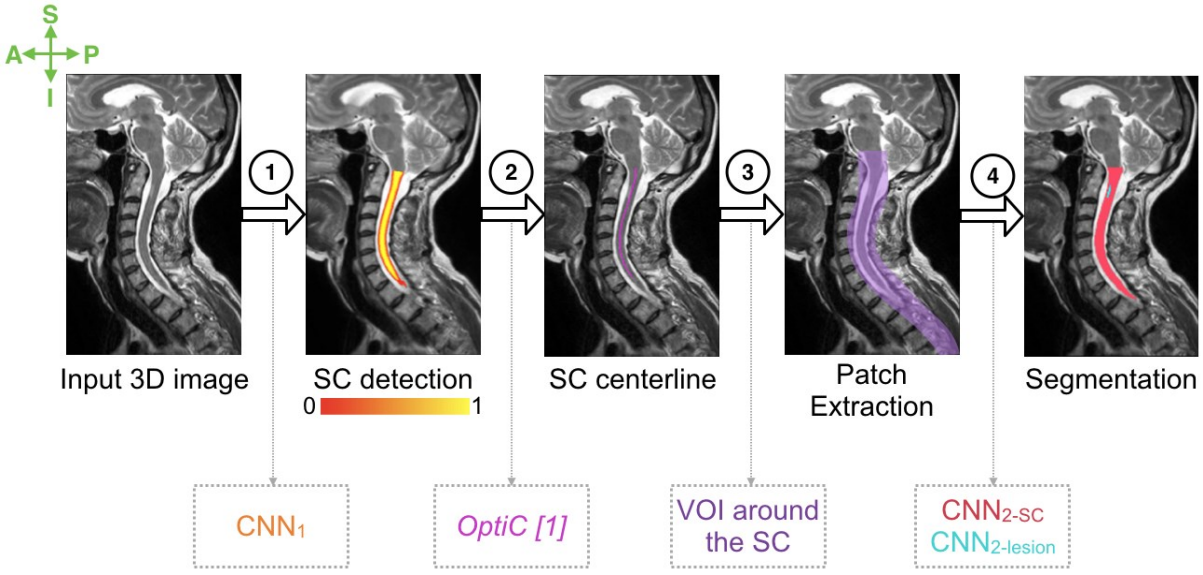


Figure 4.3: Automatic segmentation framework.

(1) detection of the spinal cord by CNN_1 outputting a heatmap of the spinal cord location, (2) computing the spinal cord centerline (pink) from the spinal cord heatmap [1]: (Gros et al., 2018), (3) extraction of 3D patches in a volume of interest surrounding the spinal cord centerline (purple), (4) segmentation of the spinal cord (red) and lesions (blue) by $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ respectively. SC: Spinal cord ; CNN: Convolutional Neural Network ; VOI: Volume of Interest ; S: Superior ; I: Inferior ; A: Anterior ; P: Posterior.

4.2.2.1 Sequential framework

In this work, automatic segmentation of the spinal cord and lesions are based on CNNs. However, CNNs can easily overfit because of two main features of our dataset: (i) the high class imbalance due to the small number of voxels labeled as positive (~ 0.34 % for spinal cord, ~ 0.01 % for lesions) and (ii) the limited number of available labeled images. To prevent overfitting, we split the learning scheme into two phases: the first phase consists of cropping the image around

the tissue of interest after which the second phase segments it. The motivation behind the sequential approach is that CNNs have been shown to learn a hierarchical representation of the provided data since the stacked layers of convolutional filters are tailored towards the desired segmentation (Christ et al., 2017; LeCun et al., 2015; Valverde et al., 2017a). The sequential framework ensures that (1) CNN_1 learns filters to discriminate between the axial patches that contain spinal cord voxels versus patches that do not, (2) while $CNN_{2\text{-}SC}$ (and $CNN_{2\text{-}lesion}$) arranges a set of filters that are specific to the spinal cord (and the lesions) from patches centered on the spinal cord centerline output by CNN_1.

4.2.2.2 Preprocessing

Preprocessing steps include resampling to 0.5 mm isotropic images (based on preliminary optimizations), re-orientation (RPI, i.e. Right-to-left, Posterior-to-anterior, Inferior-to-superior) and, finally, given the considerable range in the mean image intensities (4.7 to 420000) of our dataset, an intensity normalization algorithm (Nyúl and Udupa, 1999; Shah et al., 2011).

4.2.2.3 Spinal cord detection

4.2.2.3.1 2D patches extraction

For each volume, we extracted 2D patches (96x96) from the cross-sectional slices. We computed the mean intensity and standard deviation across all training patches to normalize all the processed patches (i.e. zero mean and unit variance), including the validation and testing patches.

4.2.2.3.2 CNN_1 architecture

CNN_1 aims at detecting the spinal cord on each cross-sectional slice of the 3D volumes (see Step 1 in **Figure 4.3**). We adapted the U-net architecture (Ronneberger et al., 2015) to use only 2 downsampling layers (rather than the 4 of the original paper) and a dilated convolution in the contracting path. Briefly, dilated convolution is a convolution with defined gaps, which provides an exponential expansion of the receptive view with a linear increase of parameters (Yu and Koltun, 2015). The motivation behind the use of dilated convolutions is to capture more contextual information (provides a broader view of the input), with fewer parameters compared to a conventional solution, which involves additional downsampling layers. Preliminary experiments led us to use a dilation rate of 3 (i.e. a gap of two pixels per input, as also illustrated in Figure 1 of (Yu and Koltun, 2015)).

4.2.2.3.3 Training strategy

Training of the CNN_I is based on the Adam optimizer (Kingma and Ba, 2014), with a learning rate of 0.0001, a batch size of 32, and 100 epochs. We applied the rectified linear activation function (Nair and Hinton, 2010) to all layers. The Dice loss (Milletari et al., 2016) was observed to be insensitive to high class imbalance (Drozdzal et al., 2018; Perone et al., 2017; Sudre et al., 2017), which encouraged us to use it as the loss function. To reduce overfitting, Batch Normalization (Ioffe and Szegedy, 2015) and Dropout (Srivastava et al., 2014), with p = 0.2, follow each convolution layer. We performed an extensive data augmentation of the training samples, including scaling, flipping, rotation and elastic deformations (Simard et al., 2003). Elastic transformations were shown to be efficient at increasing learning invariance (Dosovitskiy et al., 2014) and realistic variation in tissue (Ronneberger et al., 2015).

4.2.2.3.4 CNN_I inference and spinal cord centerline extraction

We reconstructed a volume from the patch inference of CNN_I, where values indicate the degree of confidence regarding the spinal cord location. Because CNN_I outputs a spinal cord probability mask, we calculated the Euclidean distance map from the CNN_I output to assist with spinal cord detection (see step 1 in **Figure 4.3**). We inferred the centerline from this spinal cord distance map using *OptiC* (Gros et al., 2018), a previously published fast global-curve optimization algorithm, which regularizes the centerline continuity along the Superior-to-Inferior axis (see step 2 in **Figure 4.3**).

4.2.2.4 Spinal cord and MS lesions segmentation

4.2.2.4.1 3D patches extraction

From each volume, we extracted 3D patches along the previously predicted spinal cord centerline with the following sizes: 64x64x48 for the spinal cord and 48x48x48 for MS lesions (see Step 3 in **Figure 4.3**). In preliminary experiments, we investigated different patch sizes (32x32x32, 48x48x48, 64x64x48, 96x96x48) and decided on a compromise between the class imbalance, the risk of overfitting, and the computational cost. The extracted patches overlapped with one another by 16 voxels in the Superior-to-Inferior axis. Following the same process as in **paragraph 4.2.2.3.1**, we normalized the patch intensities by centering the mean and normalizing the standard deviation.

4.2.2.4.2 $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ architecture

We trained $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ to segment the spinal cord and the MS lesions on the extracted 3D patches. The adopted network architecture draws from the 3D U-net scheme (Çiçek et al., 2016) ; however, we reduced the depth of the U-shape compared to the original paper (2 instead of 3), thus reducing the number of parameters and the amount of memory required for training.

4.2.2.4.3 Training strategy

We trained $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ using the rectified linear activation function, the Adam optimizer, the Dice loss, the Dropout (p=0.4), and the following parameters: a batch size of 4, learning rate of $5x10^{-5}$ and total number of epochs of 300. Besides flipping operations, the data augmentation procedure also includes small local erosions and dilatations of the manual lesion shapes which serve to test the confidence of the network on subjective lesion borders.

4.2.2.4.4 $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ inference

During the inference stage, $CNN_{2\text{-}SC}$ and $CNN_{2\text{-}Lesion}$ segment 3D patches extracted from a testing image with no overlap. We applied a threshold of 0.5 to the CNNs predictions before reconstructing a 3D volume (see Step 4 in **Figure 4.3**). The presented framework does not contain additional post-processing.

4.2.3 Implementation

We implemented the proposed method in the Python 2.7 language, using Keras[4] and TensorFlow[5] libraries. CNN training was carried out on a single NVIDIA Tesla P100 GPU with 16GB RAM memory. To reproduce the presented framework, a Jupyter notebook is available at [URL][6]. Moreover, the presented methods are readily available through the functions `sct_get_centerline`, `sct_deepseg_sc` and `sct_deepseg_lesion`, as part of the Spinal Cord Toolbox (SCT) (De Leener et al., 2017a). Contrary to the training which requires high computational power such as that offered by a GPU, inference (i.e. segmentation) can run in only a few minutes on a standard CPU. The models presented in this paper are freely available in SCT version v3.1.2.

[4] https://keras.io/, version 2.6.0
[5] https://www.tensorflow.org/, version 1.3.0
[6] A hyperlink will be added on a revised version of the manuscript.

4.2.4 Evaluation

For each contrast, we randomly split the dataset into three independent subsets: training (80%), validation (10%), and testing (10%). We evaluated the proposed framework on the testing dataset. To assess the robustness of the proposed method against data coming from a center unseen during the training procedure, we assured that data from two of the participating sites (n=57) was confined to the testing set.

4.2.4.1 Comparison with related works

We compared the spinal cord detection method (Steps 1 and 2 in Figure 4.3) to a recently-published study (Gros et al., 2018) that introduced a global curve optimization algorithm (*OptiC* - Step 2) but used a trained Support-Vector-Machine (SVM) algorithm to produce the spinal cord heatmap (instead of the CNN_l at Step 1). We refer to this as "*SVM+OptiC*" below.

We compared the spinal cord segmentation method to a previously-published unsupervised method, *"PropSeg"*, which is based on multi-resolution propagation of tubular deformable models (De Leener et al., 2015).

4.2.4.2 Quality metrics

4.2.4.2.1 Spinal cord detection

We evaluated the spinal cord detection by computing (i) the Mean Square Error (MSE) between the predicted and manual spinal cord centerlines, (ii) the localization rate, defined as the percentage of axial slices for which the predicted centerline was included in the manually-segmented spinal cord. We generated the manual spinal cord centerlines by computing the center of mass of each axial slice of the manual spinal cord segmentations, regularized with an approximated non-uniform rational bezier spline (NURBS), as described in (De Leener et al., 2017b).

4.2.4.2.2 Spinal cord and MS lesion segmentation

We assessed the segmentation performance with (i) the Dice Similarity Coefficient (Dice, 1945) and (ii) the relative volume difference in segmented volume (RVD, asymmetric metric) between the automatic and the manual segmentation masks. Note that, for the MS lesion segmentation task, these metrics were only computed on the MS cohort.

4.2.4.2.3 MS lesion identification

Monitoring the lesion count in the spinal cord is an important measure of disease activity, since each central nervous location where a new lesion appears would represent an entry point of the immune cells that mediate the inflammatory-demyelinating process, in other words, a breach of the blood brain-barrier. In the clinical setting, spinal cord lesion count provides complementary information to what is obtained by brain lesion monitoring in patients with MS (Healy et al., 2017). We thus computed the lesion-wise sensitivity and precision on MS data in order to assess the ability of the method to detect lesions as entities, without suffering from subjective lesion border issues. Lesion-wise sensitivity and precision were defined as:

$$sensitivity = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

where TP (true positives), FP (false positives) and FN (false negatives) are, respectively, the individual lesions correctly detected, incorrectly detected, and undetected by the automatic method. We considered a lesion as correctly detected by the automatic algorithm if it overlapped with at least one-fourth of the manual segmentation voxels.

To assess the ability of the proposed method to segment MS lesions without prior knowledge of patient status, we computed the false positive rate (FPR) on testing volumes that have been reviewed by radiologists and that did not show any lesions (i.e. subjects without an established MS diagnosis or any confirmed spinal cord lesions). We considered a volume false positive if at least one lesion was detected by the framework.

4.2.4.3 Inter-rater variability of the MS lesion segmentation

We estimated the inter-rater variability of lesion segmentation on a randomized subset of patients (n=10). For this purpose, we calculated the Dice coefficient between each rater's segmentation and a consensus reading mask, produced using "majority voting" across all the raters' labels.

4.3 Results

4.3.1 Spinal cord detection and segmentation

Figure 4.5 illustrates qualitative samples of spinal cord segmentation from the testing dataset, comparing the manual against the automatic delineation. From visual inspection, the proposed method achieved encouraging results on (i) compressed and atrophied cords (e.g., see *S5_DCM17, S5_DCM2, S25_ALS10*), (ii) slices with poor contrast between cord and surrounding structures like cerebrospinal fluid (*S16_HC1)* or MS lesions (*S15_MS24*) and (iii) images with different Superior-to-Inferior coverage, e.g. including the brain (*S4_HC15*) or thoraco-lumbar levels (*S20_MS101*).

Figure 4.4 (A.) presents the medians and interquartile ranges of the metrics evaluating the spinal cord detection and segmentation. The spinal cord detection using CNN outperformed the SVM-based method, as shown by the median MSE (interquartile range, IQR) of 1.0 (0.8) mm versus 5.5 (9.7) mm (averaging the results across all contrasts). While the two approaches produced similar results for axial data (i.e. isotropic resolution or axial orientation), improvements using the CNN approach were substantial for sagittal scans: median MSE 1.1 mm for "*CNN_l+OptiC*" versus 11.6 mm for "*SVM+OptiC*". In volumes that included part of the brain, the method accurately separated brain and spine regions in 87.0 % of cases (i.e. between the top of C1 and the ponto-medullary junction). The median MSE was largely improved by resorting to the curve optimization algorithm, especially on DCM patients, as it considerably decreased from 24.04 mm (CNN_l output, step 1 in **Figure 4.3**) to 1.14 mm ("*CNN_l+OptiC*" output, step 2 in **Figure 4.3**).

Regarding the segmentation task, the proposed spinal cord framework achieved superior results compared to *PropSeg*, with a median (IQR) Dice of 94.6 (4.6) versus 87.9 (18.3) %. In particular, the proposed method outperformed *PropSeg* in patients with severe cord atrophy in terms of (i) Dice: 92.9 % vs. 82.0 % and (ii) relative volume difference: -3.6 % vs. +13.3 %. Axial scans were more precisely segmented than sagittal scans (median RVD: -4.8 % versus -10.2 %). The proposed framework was robust to MS-related pathology since the automatic segmentation yielded similar results between controls and MS subjects (median Dice: 95.2 % versus 94.1 %). The model generalized well to data from a center unseen during the training (median Dice of 93.3 %). For a typical T_2-weighted acquisition (matrix size: 384x384x52,

resolution: 1 mm isotropic), the computation time on an iMac (i7 4-cores 3.4 GHz 8Gb RAM), including reading and writing tasks, was 1 minute 55 seconds for the proposed method versus 32 seconds for *PropSeg*.

4.3.2 MS lesion segmentation

Figure 4.6 depicts several qualitative examples of MS lesion segmentations (both manual and automatic) from the testing dataset. The main divergence between manual and automatic segmentations were located near normal-appearing structures (e.g. cerebrospinal fluid, grey matter) where the partial volume effect challenged tissue delineation (e.g. samples *S1_SPMS9, S2_RRMS5*). However, a visual inspection of the results shows that the network successfully learned the pattern of the normal-appearing grey matter despite its confounding intensities with MS lesions (e.g. samples *S7_RRMS14*). Instances where the automatic method correctly detected small lesions as well as lesions in atrophied cord are also shown in **Figure 4.6** (see *S1_RRMS17, S2_CIS1, S8_PPMS10*). Although the Dice metric is widely used for medical image segmentation, it should be noted that it has a larger dynamic sensitivity to small vs. large objects (see *S2_CIS1, S3_RRMS7)*.

Figure 4.4 (B.) shows the medians and interquartile ranges of the metrics evaluating the automatic MS lesion segmentation. When pooling T_2-w and T_{2^*}-w, the automatic segmentation method reached a median (IQR) Dice of 60.0 (21.4) %. While this result might appear weak, it should be seen in light of the inter-rater study, where the raters achieved a median Dice of 60.7 % compared to 56.8 % for the automatic method. In terms of volumetric considerations, the automatic method provided satisfactory results, exhibiting a reasonable under-segmentation rate (median RVD: -14.5 %). Regarding the lesion-wise detectability, the automatic method yielded a low number of false negative (median sensitivity: 83.3 %) and false positive (median precision: 76.9 %) lesion labels. The method's sensitivity was slightly better in detecting lesions on sagittal vs. axial scans (median sensitivity: 100.0 % versus 75.0 %). When confronted with data from sites excluded from the training dataset, the method provided similar results as other sites (median sensitivity: 100.0 %, median Dice: 57.0 %). Finally, when encountering a subject without an established MS diagnosis or any confirmed spinal cord lesions, the automatic method produced a fairly high false positive rate by detecting at least one lesion on 32.0 % of the cases.

Figure 4.7 compares the raters and automatic MS lesion segmentation on 10 testing subjects. An inter-rater variability was observed: the Dice results varied by 85.0 % among the raters for subject 004, and by 21.0 % for subject 008 (see **Figure 4.7 A.**). The disagreements between raters mainly occurred on the borders of the lesions, in particular, the lesion extension within the grey matter area on T_2^*-w images (see **Figure 4.7 B.**). The average time for manually segmenting lesions in one subject (two volumes per subject) was 18.7 minutes vs. 3.6 minutes using the automatic method (iMac i7 4-cores 3.4 GHz 8Gb RAM).

A. Spinal Cord Detection and Segmentation

	Detection				Segmentation			
	Mean Square Error		Localization Rate		Dice Coefficient		Relative Volume Difference	
	Best: 0.0mm		*Best: 100.0%*		*Best: 100.0%*		*Best: 0.0%*	
	SVM + OptiC	*CNN_1*	*SVM+OptiC*	CNN_1	*PropSeg*	*CNN_1 +$CNN_{2\text{-}SC}$*	*PropSeg*	*CNN_1+$CNN_{2\text{-}SC}$*
T_1-w Dataset	11.1 (11.8)	0.9 (0.5)	33.3 (48.9)	100. (0.)	92.0 (13.5)	95.9 (1.5)	-4.4 (11.1)	-0.3 (5.7)
T_2-w Dataset	9.1 (12.8)	1.0 (0.9)	100. (33.3)	99.7 (4.2)	83.2 (18.6)	92.4 (5.1)	7.0 (26.8)	-0.2 (6.5)
T_2^*-w Dataset	0.9 (0.3)	1.0 (0.6)	100. (0.)	100. (0.)	94.1 (15.7)	95.5 (2.8)	4.3 (32.8)	-3.5 (9.8)

B. MS Lesion Segmentation

	Dice Coefficient	Relative Volume Difference	Sensitivity	Precision
	Best: 100.0%	*Best: 0.0%*	*Best: 100.0%*	*Best: 100.0%*
T_2-w Dataset	57.6 (22.4)	-17.3 (61.3)	90.0 (33.3)	66.7 (58.3)
T_2^*-w Dataset	60.4 (25.0)	-4.5 (74.9)	75.0 (47.2)	100.0 (38.4)

Figure 4.4: Quantitative results of the fully-automatic framework.

Results on the testing dataset of the automatic (A) detection and segmentation of the spinal cord, including a comparison with previously published methods: "*SVM + OptiC*" (Gros et al., 2018) and "PropSeg" (De Leener et al., 2015) (B) segmentation of the MS lesions. Data indicate median (interquartile range) for each contrast dataset: T_1-weighted, T_2-weighted and T_2^*-weighted.

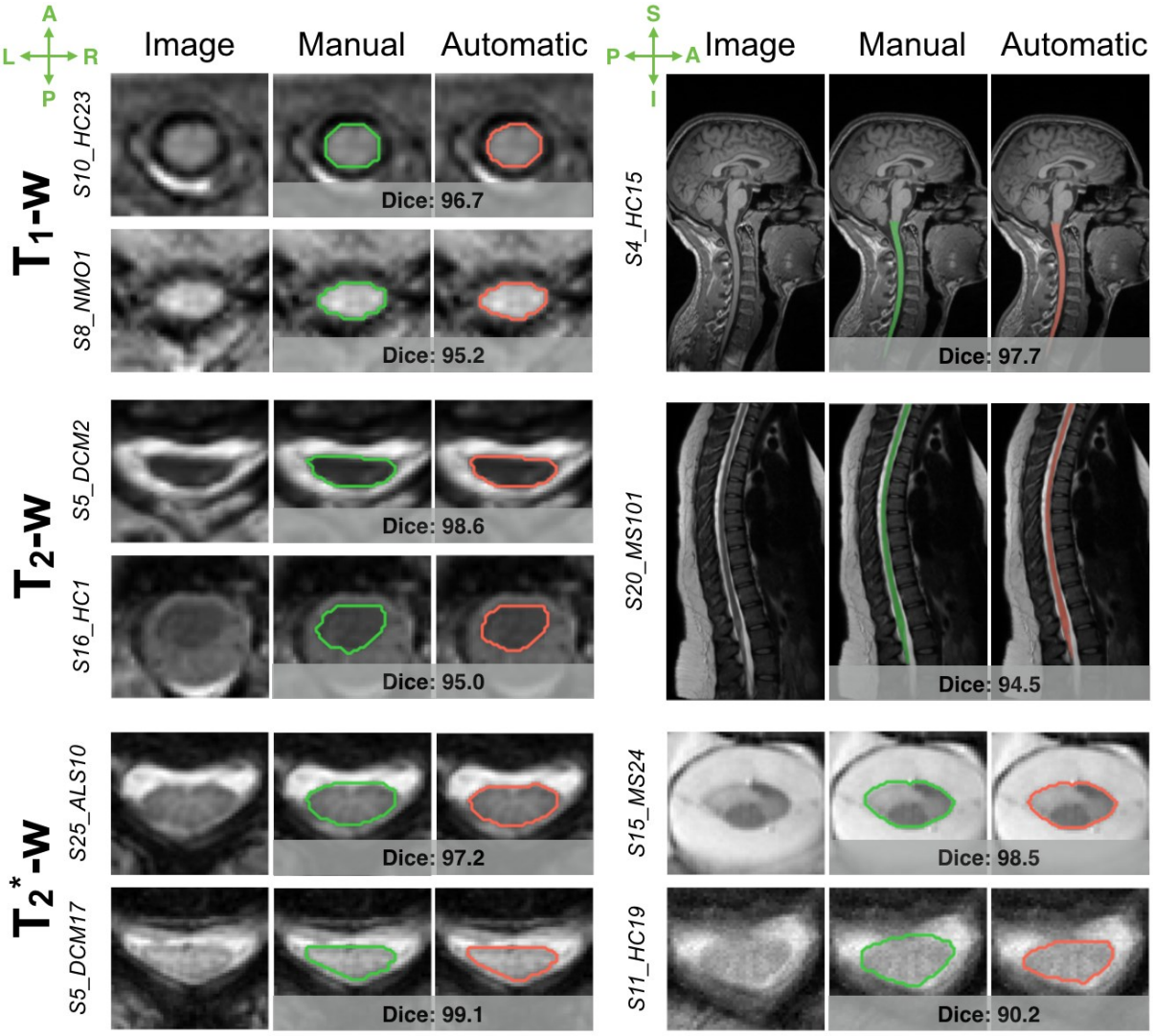


Figure 4.5: Quantitative results of the fully-automatic spinal cord segmentation.

Examples of automatic spinal cord segmentations on T_1-w (top), T_2-w (middle) and T_2^*-w (bottom) MRI data. This includes a comparison between manual (green) and automatic (red) delineations, with Dice coefficient indicated just below each comparison. Abbreviations: A: Anterior ; P: Posterior ; L: Left ; R: Right ; I: Inferior ; S: Superior ; Auto.: Automatic ; HC: healthy controls ; MS: multiple sclerosis ; DCM: degenerative cervical myelopathy ; NMO: neuromyelitis optica ; ALS: amyotrophic lateral sclerosis. Note that the depicted samples represent a variety of subjects in terms of clinical status, and were scanned at different sites, identified by their ID (e.g. S10_HC23 is the ID of HC subject #23, from the site #10).

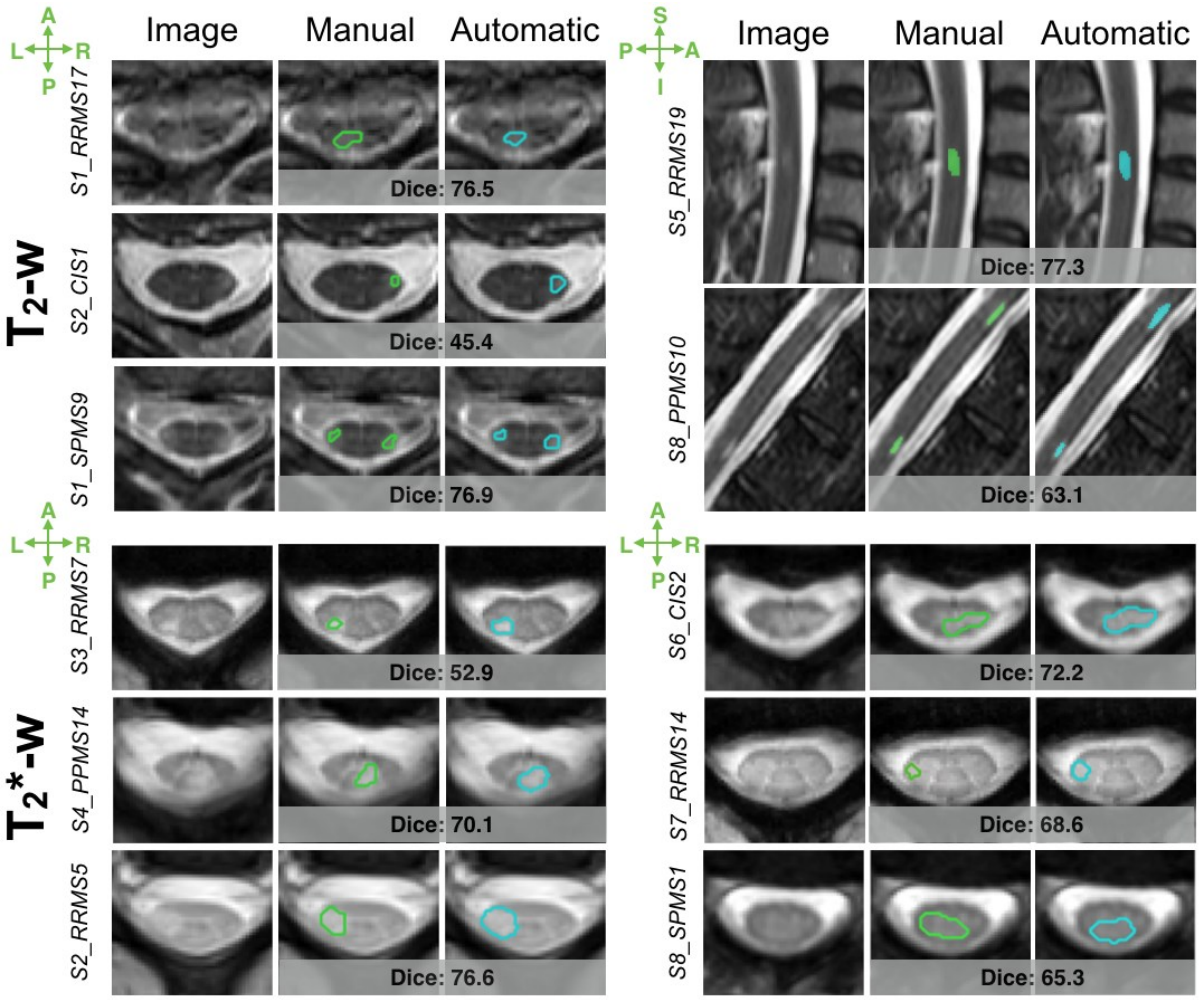
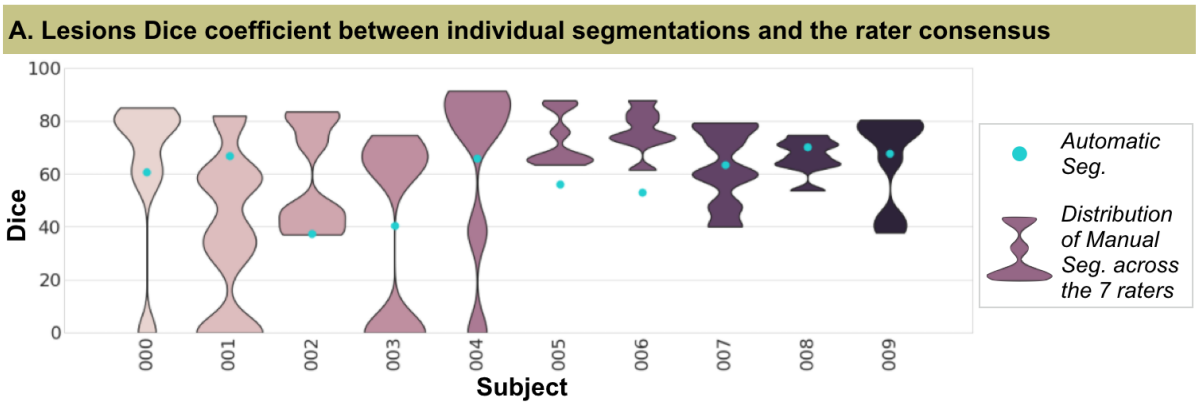
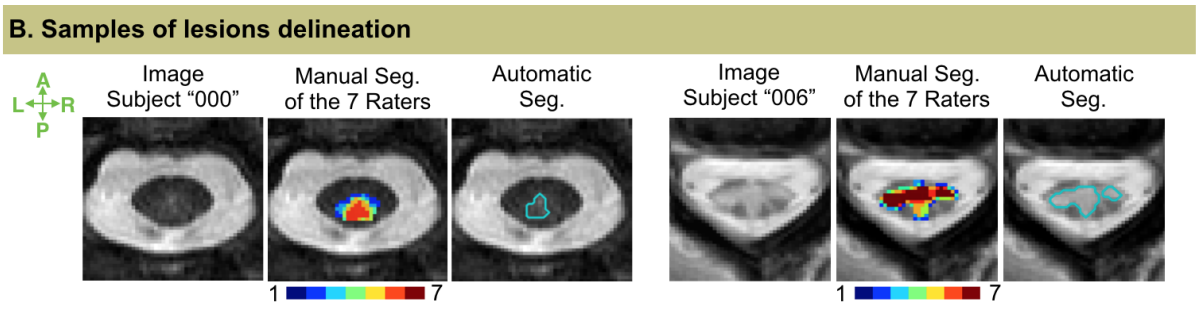


Figure 4.6: Quantitative results of the fully-automatic intramedullary lesions segmentation.

Examples of automatic lesion segmentations on Axial T_2-w (top left), Axial T_2*-w (bottom) and Sagittal T2-w (top, right) MRI data. This includes a comparison between manual (green) and automatic (blue) delineations, with Dice coefficients indicated just below each comparison. Note that the depicted samples were scanned at different sites, identified by their ID (e.g. S1_RRMS17 is the ID of subject #17 from site #1 with relapsing-remitting multiple sclerosis).





C. Lesions segmentation time

Subject	Rater #1	Rater #2	Rater #3	Rater #4	Rater #5	Rater #6	Rater #7	Auto. Seg.
Mean seg. time per subject [min]	Not Available	17.5	24.5	11.2	6.5	15.5	36.8	3.6

Figure 4.7: Intra-rater study.

Comparison between raters and automatic MS lesion segmentation on 10 subjects from the "testing" group. (A.) shows the Dice coefficient (best value: 100) computed between the rater consensus (majority voting) and each individual rater segmentation (purple distributions) as well as the automatic method (blue dot). (B.) depicts axial cross-sectional samples with the manual segmentation of the raters and the automatic delineation (blue). The consensus between raters vary from "low agreements" (in blues, mainly on the borders) to "strong agreement" (in reds, mainly on the cores). (C.) presents the segmentation time, averaged across subjects, for each rater and the automatic segmentation (iMac i7 4-cores 3.4 GHz 8Gb RAM). Abbreviations: Seg.: Segmentation ; A: Anterior ; P: Posterior ; L: Left ; R: Right ; I: Inferior ; S: Superior ; Auto.: Automatic.

4.4 Discussion

We introduced a robust method to segment the spinal cord and MS lesions contained within it. The proposed framework is based on a sequence of two CNNs, trained individually to tailor a set of specific filters for each targeted structure. The first network is trained to detect the spinal cord within the 3D volume, so that the volume investigated by the second network is restricted to a close neighborhood of the target structures: the spinal cord and MS lesions. Furthermore, the framework has been designed to handle the heterogeneity of image acquisition features. Evaluation was performed on a large multi-center cohort including participants with various clinical conditions as well as healthy controls. The developed tools `sct_get_centerline`, `sct_deepseg_sc` and `sct_deepseg_lesion` are freely available as part of SCT (De Leener et al., 2017a).

4.4.1 Spinal cord detection

Robustly localizing the spinal cord centerline on MRI data is a key step for automating spinal cord segmentation (De Leener et al., 2015; Horsfield et al., 2010) and template registration (De Leener et al., 2018; Stroman et al., 2008). The proposed method works in two steps: (i) recognition by a CNN of the spinal cord pattern on axial slices and (ii) regularization of the spinal cord centerline continuity along the Superior-to-Inferior direction using a global curve optimization algorithm (Gros et al., 2018). Although the spinal cord pattern was well identified by CNN_1 in the first step, resorting to the curve regularization (step ii) was important for ensuring centerline consistency. This was especially true for patients with spinal cord atrophy, for whom the contrast between the cerebrospinal fluid and the spinal cord was frequently very low in large sections of the cord. Having produced detections of similar accuracy for axial and sagittal scans, this approach demonstrated its robustness to image resolution, particularly when compared to its predecessor (Gros et al., 2018). In addition, the new method can be used to separate spine and brain sections, which are regularly covered during cervical scans.

4.4.1.1 Limitations

The introduction of a detection step prior to the segmentation module was motivated by the high class imbalance (proportion of spinal cord and/or lesion compared to the rest of the volume) and the large heterogeneity of image features (contrast, field of view, etc.). However, the disadvantage of the sequential approach is that the segmentation framework is sensitive to the

quality of the detection module. Fortunately, though, the high performance of the spinal cord detection (median MSE of 1. mm) is reliable enough to be cascaded by another CNN. When scans incorporated the brain, 13% of the spinal cord centerlines extended above the ponto-medullary junction, but without impacting the consecutive cord segmentation.

4.4.1.2 Perspectives

Besides the three MR contrasts investigated in this study (T_1-, T_2- and T_2^*-weighted), we plan to cover other commonly-used sequences, such as diffusion-weighted scans and T_2^*-weighted echo-planar imaging (typically used for fMRI studies) and to make the additional trained models available in SCT. Apart from segmentation purposes, the centerline spatial information could guide an automatic tool for identification of the vertebral discs along the spinal canal (Ullmann et al., 2014), provide spinal cord curvature information for studying the biomechanics of the spine and planning surgery (Gervais et al., 2012; Little et al., 2016), or be used for localized shimming (Topfer et al., 2018, 2016; Vannesjo et al., 2017).

4.4.2 Spinal cord segmentation

Spinal cord segmentation has important clinical value for measuring cord atrophy in MS patients (Dupuy et al., 2016; Kearney et al., 2014; Losseff et al., 1996; Lundell et al., 2017; Rocca et al., 2013, 2011; Singhal et al., 2017). Besides MS pathology, spinal cord segmentation could provide a valuable quantitative assessment of spinal cord morphometry in the healthy population (Fradet et al., 2014; Papinutto et al., 2015) or be used as a biomarker for other spinal cord diseases (Martin et al., 2017; Nakamura et al., 2008; Paquin et al., 2018). We proposed an automatic method to segment the spinal cord without prior knowledge, and validated the method against manual segmentation on a multi-center clinical dataset involving a variety of pathologies. We also compared this method to the previously published *PropSeg* method (De Leener et al., 2015). The proposed method achieved better results than *PropSeg* in terms of Dice and relative volume difference, especially in patients with severe cord compression. The *PropSeg* deformable model tends to leak out of the cord when cerebrospinal fluid/spinal cord contrast is low, while the $CNN_{2\text{-}SC}$ could avoid this by incorporating larger spatial information. The framework performed equally across 3 different MR contrasts (T_1-, T_2- and T_2^*-weighted) and was robust to field of view heterogeneity, which is a necessary feature to ensure its use in multi-site data analysis pipelines. When presenting our model with data from new sites, performance was similar to when the data

came from the original sites (i.e. sites included in the supervised learning). The ability of our model to generalize is likely due to the large training dataset, mostly composed of 'real-world' clinical data and spanning a broad diversity of scanning platform and acquisition parameters.

4.4.2.1 Limitations

Though the deformable model of *PropSeg* could be adjusted in cases of segmentation failure (e.g. alter the radius of the SC, or conditions of the deformation), there is less room with the CNN-based approach for changing input parameters during inference. Moreover, the presented method is slower than *PropSeg*, mainly due to the use of 3D convolutions (see **paragraph 4.4.4.2**). It is, however, important to note that the evaluation was biased in favour of *PropSeg*, since most of the manual spinal cord delineations were produced by correcting the mask previously generated by *PropSeg*.

4.4.2.2 Perspectives

To improve image quality and reduce the variability across sites, preliminary experiments explored the impact of advanced preprocessing techniques, such as denoising (Coupe et al., 2008) and bias field correction (Tustison et al., 2010). Finding a set of generic preprocessing hyper-parameters that works for every dataset is challenging. Preprocessing, fine-tuned for a specific and homogeneous dataset, however, could improve the segmentation. Along with the spinal cord, the automatic segmentation of the cerebrospinal fluid could also provide a measure of the spinal canal volume for normalizing cord volumes across people of different sizes, analogous to brain parenchymal fraction or brain to intra-cranial capacity ratio. Finally, the scan-rescan reproducibility of the proposed segmentation method will be the subject of future investigations.

4.4.3 MS lesion segmentation

Automating spinal cord MS lesion segmentation provides an efficient solution to evaluate large data sets for lesion burden analyses. A thorough search of the relevant literature did not yield available related work. Results of the automatic segmentation were similar to the inter-rater results, with the advantage of higher efficiency and reproducibility (i.e. it will always produce the same segmentation for the same image). The high sensitivity and precision in detecting lesions are encouraging since these markers are particularly relevant to clinical studies (Thompson et al., 2018). While the Dice scores were quite low (median: 60.0 %), it should be noted that this metric is highly sensitive to the total lesion load and lesion sizes (Guizard et al., 2015; Harmouche et al.,

2015; Styner et al., 2008). The median Dice of 60.7 % between each rater and the consensus reading illustrates that point well, which is in line with recent inter-rater variability results obtained on brain lesions: 63% (Carass et al., 2017) and 66% (Egger et al., 2017). This is why we also computed complementary metrics which are less subjective to lesion borders (e.g. object detection precision and sensitivity) (Geremia et al., 2011; Harmouche et al., 2015; Lladó et al., 2012; Styner et al., 2008; Valverde et al., 2017a). The relative volume difference was also used since the total lesion load is frequently used as a clinical biomarker. The proposed method has shown to generalize well to data exhibiting features which were absent in the training data. Multiple instances of prominent artifacts and poor image quality were found in the dataset. However, these cases were nevertheless retained for the study, considering that they are part of the clinical reality.

4.4.3.1 Limitations

The main detection failures (false positives and/or false negatives) were observed with small lesions (< 50 mm^3). The results of the automatic method, as well as the raters' assessments, hinted at variable levels of detectability across sites. Variations in sequences and image contrast might account for the observed differences in performance. Recent initiatives to standardize spinal cord MRI protocols will likely help to reduce such variability in the future (Alley et al., 2018). In volumes where no lesions were detected by raters, the false positive rate was high (32.0 %). These false segmentations were generally very small (median individual volume: 12.5 mm^3) and likely due to the partial volume effect between the cord and cerebrospinal fluid. Using data acquired with isotropic resolution (to minimize partial volume effect in one direction) and/or CNN architectures based on multimodal data (Havaei et al., 2016) would likely reduce the false positive rate and will be investigated in future studies (see also the next section below).

Lesion borders can often be diffuse, so that defining an "edge" can be somewhat arbitrary and highly subjective in these cases. As a result, lesion borders are frequently the site of disagreement between manual and automatic delineations, as well as among raters. This motivated our implementation of a data augmentation module to prompt the model to be less confident of the lesion border prediction (random and local erosion/dilation of the lesion masks during the training). Its specific effect on the segmentation performance will be validated in future work.

4.4.3.2 Perspectives

In this work, MS lesion segmentation was achieved by processing each 3D scan independently, which is arguably a non-optimal use of the different available contrasts. In clinical settings however, it is not uncommon to have more than one acquisition covering the same region. Future work could consider recent advances in domain-adaptation (Ghafoorian et al., 2017; Valindria et al., 2018) to overcome variations in imaging protocols. Indeed, a combination of the information from different MR contrasts should help the identification of very small lesions while reducing the number of false positives. Considering that image labelling is time consuming and tedious, semi-supervised learning approaches should be explored to take advantage of the wide number of available unlabeled data (Baur et al., 2017). Another exploratory avenue could be to explore the patterns that have been automatically learned by the CNN (see **Figure 4.8**), as suggested by a recent study on brain lesions (Kamnitsas et al., 2017). For example, we were surprised by the ability of the network to distinguish lesions in the normal-appearing grey matter on T_2*-weighted scans, suggesting that the pattern of the healthy grey matter has been self-learnt. This observation could suggest that great potential lies in the combination of the CNN discriminative ability and clinical knowledge, such as spatial priors for cervical lesions. This is in line with previous segmentation work, where performance of traditional classifiers was significantly improved by incorporation of tissue priors (Harmouche et al., 2015; Shiee et al., 2010; Van Leemput et al., 1999). It would thus be interesting to investigate ways for encoding such available prior information into the network's feature space, so that clinical knowledge could direct the network towards the optimal solution. This could indeed drastically simplify the optimization problem and mitigate false positive detections.

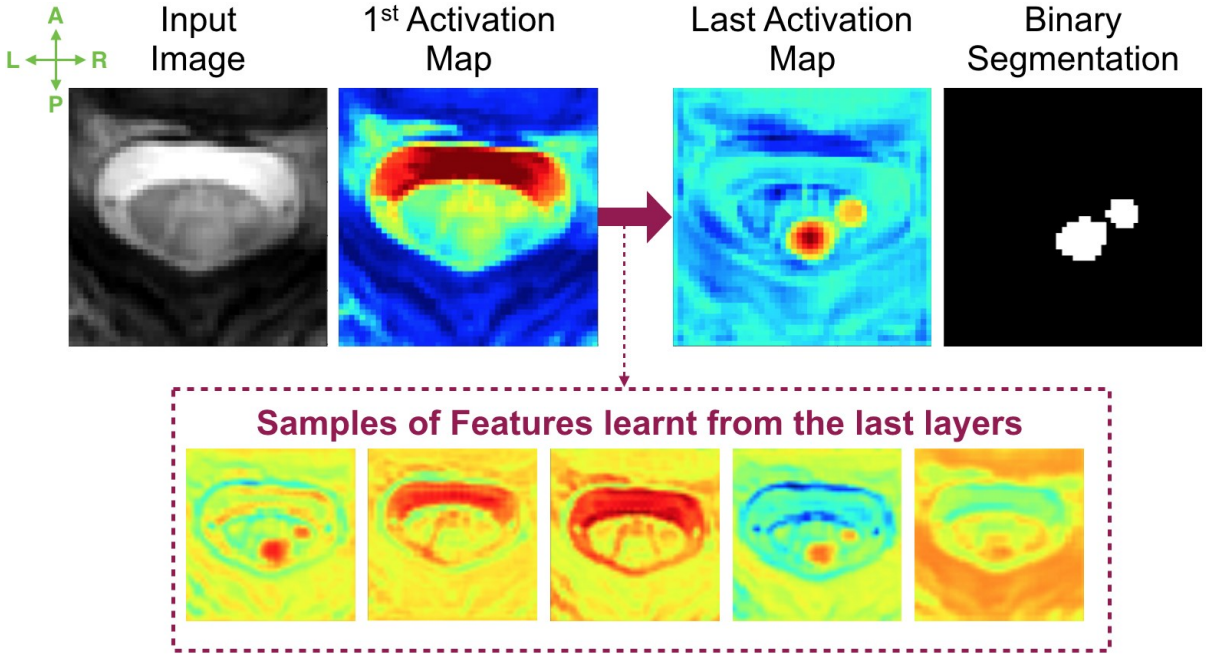


Figure 4.8: Feature maps.

Visualization of feature map instances, learnt by different layers of the $CNN_{2\text{-}Lesion}$, applied to an input image (left) leading to a binary segmentation (right).

The performance of our method is hampered when it comes to segmenting small lesions. To mitigate this challenge, recent work suggest adding a secondary training stage (Cireşan et al., 2013; Havaei et al., 2015; Valverde et al., 2017a). Briefly, the second stage of Valverde et al. was trained with challenging samples by including the false positives of the first network in its learning. Other perspectives directly related to the CNNs training will be presented in the following section.

4.4.4 CNNs Training

Due to the large heterogeneity in MRI contrast (see Figure 2), images were distributed among three MRI contrast datasets, for both the training and inference of the CNNs: (i) "T_1-weighted like" (i.e. dark cerebrospinal fluid / light cord), (ii) "T_2-weighted like" (i.e. light cerebrospinal fluid / dark cord / grey matter not visible), (iii) "T_2^*-weighted like" (i.e. light cerebrospinal fluid / dark cord / grey matter visible). The performance of the framework was consistent when trained with the 3 different MR contrast datasets, which highlights its robustness to different training conditions.

4.4.4.1 Class imbalance

An important challenge to the design of automated MS lesion segmentation methods is the extremely unbalanced nature of the data. In this work, this issue of class imbalance was mitigated by using the Dice loss, and by restricting the search around the spinal cord centerline facilitated by the first CNN.

In preliminary experiments, we explored the benefit of under-sampling the negative class during the training to address the massive class imbalance. While it significantly facilitated the training convergence, it biased the classifier towards the positive class and may have resulted in a drastic increase in false positive detections. More complex sampling schemes (Havaei et al., 2015; Jesson et al., 2017; Valverde et al., 2017a), successfully employed in medical image segmentation or detection tasks, could be investigated for spinal cord applications.

Moreover, in exploratory experiments, we also tested various loss functions specifically proposed to mitigate the class imbalance issues: the weighted cross-entropy (Ronneberger et al., 2015), the Dice (Milletari et al., 2016) and the "sensitivity - specificity" (Brosch et al., 2015) loss functions. Although the Dice loss caused narrow boundaries of confidence intervals at the edge, it yielded better results. In the future, other loss functions fashioned to handle highly unbalanced datasets could be tested, such as the Focal Loss (Lin et al., 2017) or the Generalised Dice overlap (Sudre et al., 2017).

4.4.4.2 3D spatial information

Prior experiments also explored the use of 3D instead of 2D patches compared to the commonly used 2D patches, as they were preferred in recent work on biomedical volumes (Çiçek et al., 2016; Kamnitsas et al., 2017; Milletari et al., 2016). While 3D patches provide more context-rich information, 3D CNNs have more parameters and thus memory and computational constraints.

For the spinal cord detection step, 2D patches were used to localize the position of the cord. Two-dimensional axial patches were adopted here for the sake of computational simplicity, considering that 3D patches did not yield substantial improvements. The use of 2D dilated convolutions might account for the accurate detections. Indeed, by increasing the receptive fields, dilated convolutions benefit from a broader spatial context for detecting sparse structures, while maintaining a relatively low number of parameters to optimize.

In most cases, the spinal cord segmentation quantitative results were similar whether 2D or 3D patches were used. However, in the cases with exceptional lesion load and severe atrophy, the incorporation of 3D contextual information showed noteworthy improvements, which consequently motivated the adoption of 3D patches. As mentioned before, the use of 3D convolutions caused a drastic increase of memory consumption, computational cost and training time. Further studies could investigate solutions to reduce the memory consumption (inherent to the use of 3D convolutions), such as the Reversible Residual Network architecture (Gomez et al., 2017) or multi-stream architectures (Prasoon et al., 2013). Furthermore, future work could explore the benefit of 3D dense conditional random fields (Christ et al., 2016; Krähenbühl and Koltun, 2011; Zheng et al., 2015) to incorporate 3D context instead of using 3D convolutions.

4.5 Conclusion

We presented an original automated spinal cord and MS lesion segmentation method, based on a sequence of two convolutional neural networks. Spinal cord segmentation results outperformed a state-of-the-art method on a multi-center and highly heterogeneous clinical dataset. Lesion segmentation results were generally within the range of manual segmentations, although the false positive rate warrants further investigations. The presented automatic methods are open-source and readily accessible in SCT.

DECLARATION OF INTEREST

Charley Gros, Benjamin De Leener, Josefina Maranzano, Dominique Eden, Atef Badji, Govind Nair, Tobias Granberg, Hugh Kearney, Ferran Prados, Russell Ouellette, Daniel S. Reich, Pierre Labauge, Leszek Stawiarz, Anne Kerbrat, Elise Bannier, Shahamat Tauhid and Julien Cohen-Adad have no relevant financial interests to disclose.

Prof. Filippi is Editor-in-Chief of the Journal of Neurology; received compensation for consulting services and/or speaking activities from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries; and receives research support from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and ARiSLA (Fondazione Italiana di Ricerca per la SLA).

Jan Hillert has received honoraria for serving on advisory boards for Biogen, Sanofi-Genzyme and Novartis; and speaker's fees from Biogen, Novartis, Merck-Serono, Bayer-Schering, Teva and Sanofi-Genzyme.; and has served as P.I. for projects or received unrestricted research support from Biogen Idec, Merck-Serono, TEVA, Sanofi-Genzyme and Bayer-Schering.

M.A. Rocca received speaker honoraria from Biogen Idec, Novartis, Genzyme, Sanofi-Aventis, Teva and Merck Serono and receives research support from the Italian Ministry of Health and Fondazione Italiana Sclerosi Multipla.

Dr. S. Narayanan reports personal fees from NeuroRx Research, a speaker's honorarium from Novartis Canada, and grants from the Canadian Institutes of Health Research, unrelated to the submitted work.

P. Valsasina received speaker honoraria from Biogen Idec, Novartis and ExceMED.

Donald G. McLaren and Vincent Auclair are currently employees of Biospective, Inc.

Rohit Bakshi has received consulting fees from Bayer, EMD Serono, Genentech, Guerbet, Sanofi-Genzyme, and Shire and research support from EMD Serono and Sanofi-Genzyme.

Jean Pelletier received speaker honoraria from Biogen, Roche, Genzyme, Novartis, and research supports from the French Ministry of Health and ARSEP.

REFERENCES

Alley, S., Gilbert, G., Gandini Wheeler-Kingshott, C.A., Samson, R.S., Grussu, F., Martin, A.R., Bannier, E., Callot, V., Smith, S.A., Xu, S., Dewey, B., Weber, K.A., Parrish, T.B., McLaren, D., Barker, G.J., Papinutto, N., Seif, M., Freund, P., Barry, R.L., By, S., Narayanan, S., Cohen-Adad, J., 2018. Consensus acquisition protocol for quantitative MRI of the cervical spinal cord at 3T, in: Proceedings of the 26th Annual Meeting of ISMRM. Presented at the ISMRM.

Arrambide, G., Rovira, A., Sastre-Garriga, J., Tur, C., Castilló, J., Río, J., Vidal-Jordana, A., Galán, I., Rodríguez-Acevedo, B., Midaglia, L., Nos, C., Mulero, P., Arévalo, M.J., Comabella, M., Huerga, E., Auger, C., Montalban, X., Tintore, M., 2018. Spinal cord lesions: A modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor. Mult. Scler. 24, 301–312.

Bakshi, R., Dandamudi, V.S.R., Neema, M., De, C., Bermel, R.A., 2005. Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. J. Neuroimaging 15, 30S–45S.

Baur, C., Albarqouni, S., Navab, N., 2017. Semi-supervised Deep Learning for Fully Convolutional Networks, in: Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. Springer International Publishing, pp. 311–319.

Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsee, A., Tam, R., 2015. Deep Convolutional Encoder Networks for Multiple Sclerosis Lesion Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, pp. 3–11.

Buda, M., Maki, A., Mazurowski, M.A., 2017. A systematic study of the class imbalance problem in convolutional neural networks. arXiv preprint arXiv:1710. 05381.

Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C.H., Jorge Cardoso, M., Cawley, N., Ciccarelli, O., Wheeler-Kingshott, C.A.M., Ourselin, S., Catanese, L., Deshpande, H., Maurel, P., Commowick, O., Barillot, C., Tomas-Fernandez, X., Warfield, S.K., Vaidya, S., Chunduru, A., Muthuganapathy, R., Krishnamurthi, G., Jesson, A., Arbel, T., Maier, O., Handels, H., Iheme, L.O., Unay, D., Jain, S., Sima, D.M., Smeets, D., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Bazin, P.-L., Calabresi, P.A., Crainiceanu, C.M., Ellingsen, L.M., Reich, D.S., Prince, J.L., Pham, D.L., 2017. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. Neuroimage 148, 77–102.

Carbonell-Caballero, J., Manjón, J.V., Martí-Bonmatí, L., Olalla, J.R., Casanova, B., de la Iglesia-Vayá, M., Coret, F., Robles, M., 2006. Accurate quantification methods to evaluate cervical cord atrophy in multiple sclerosis patients. MAGMA 19, 237–246.

Chen, M., Carass, A., Oh, J., Nair, G., Pham, D.L., Reich, D.S., Prince, J.L., 2013. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. Neuroimage 83, 1051–1062.

Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M., Sommer, W.H., Ahmadi, S.-A., Menze, B.H., 2016. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 415–423.

Christ, P.F., Ettlinger, F., Grün, F., Elshaera, M.E.A., Lipkova, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Hofmann, F., Anastasi, M.D., Ahmadi, S.-A., Kaissis, G., Holch, J., Sommer, W., Braren, R., Heinemann, V., Menze, B., 2017. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. arXiv [cs.CV].

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 424–432.

Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2013. Mitosis detection in breast cancer histology images with deep neural networks. Med. Image Comput. Comput. Assist. Interv. 16, 411–418.

Cohen, A.B., Neema, M., Arora, A., Dell'oglio, E., Benedict, R.H.B., Tauhid, S., Goldberg-Zimring, D., Chavarro-Nieto, C., Ceccarelli, A., Klein, J.P., Stankiewicz, J.M., Houtchens, M.K., Buckle, G.J., Alsop, D.C., Guttmann, C.R.G., Bakshi, R., 2012. The relationships among MRI-defined spinal cord involvement, brain involvement, and disability in multiple sclerosis. J. Neuroimaging 22, 122–128.

Compston, A., Coles, A., 2002. Multiple sclerosis. Lancet 359, 1221–1231.

De Leener, B., Cohen-Adad, J., Kadoury, S., 2015. Automatic Segmentation of the Spinal Cord and Spinal Canal Coupled With Vertebral Labeling. IEEE Trans. Med. Imaging 34, 1705–1718.

De Leener, B., Fonov, V.S., Collins, D.L., Callot, V., Stikov, N., Cohen-Adad, J., 2018. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. Neuroimage 165, 170–179.

De Leener, B., Kadoury, S., Cohen-Adad, J., 2014. Robust, accurate and fast automatic segmentation of the spinal cord. Neuroimage 98, 528–536.

De Leener, B., Lévy, S., Dupont, S.M., Fonov, V.S., Stikov, N., Louis Collins, D., Callot, V., Cohen-Adad, J., 2017a. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 145, 24–43.

De Leener, B., Mangeat, G., Dupont, S., Martin, A.R., Callot, V., Stikov, N., Fehlings, M.G., Cohen-Adad, J., 2017b. Topologically preserving straightening of spinal cord MRI. J. Magn. Reson. Imaging 46, 1209–1219.

De Leener, B., Taso, M., Cohen-Adad, J., Callot, V., 2016. Segmentation of the human spinal cord. MAGMA 29, 125–153.

Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. Ecology 26, 297–302.

Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T., 2014. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp. 766–774.

Drozdzal, M., Chartrand, G., Vorontsov, E., Shakeri, M., Di Jorio, L., Tang, A., Romero, A., Bengio, Y., Pal, C., Kadoury, S., 2018. Learning normalized inputs for iterative estimation in medical image segmentation. Med. Image Anal. 44, 1–13.

Dupuy, S.L., Khalid, F., Healy, B.C., Bakshi, S., Neema, M., Tauhid, S., Bakshi, R., 2016. The effect of intramuscular interferon beta-1a on spinal cord volume in relapsing-remitting multiple sclerosis. BMC Med. Imaging 16, 56.

Egger, C., Opfer, R., Wang, C., Kepp, T., Sormani, M.P., Spies, L., Barnett, M., Schippling, S., 2017. MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? Neuroimage Clin 13, 264–270.

Filippi, M., Rocca, M.A., 2007. Conventional MRI in multiple sclerosis. J. Neuroimaging 17 Suppl 1, 3S–9S.

Fradet, L., Arnoux, P.-J., Ranjeva, J.-P., Petit, Y., Callot, V., 2014. Morphometrics of the entire human spinal cord and spinal canal measured from in vivo high-resolution anatomical magnetic resonance imaging. Spine 39, E262–9.

García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L., 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med. Image Anal. 17, 1–18.

Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage 57, 378–390.

Gervais, J., Périé, D., Parent, S., Labelle, H., Aubin, C.-E., 2012. MRI signal distribution within the intervertebral disc as a biomarker of adolescent idiopathic scoliosis and spondylolisthesis. BMC Musculoskelet. Disord. 13, 239.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.-E., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., Wells, W.M., III, 2017. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. arXiv [cs.CV].

Gomez, A.N., Ren, M., Urtasun, R., Grosse, R.B., 2017. The Reversible Residual Network: Backpropagation Without Storing Activations, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 2214–2224.

Gros, C., De Leener, B., Dupont, S.M., Martin, A.R., Fehlings, M.G., Bakshi, R., Tummala, S., Auclair, V., McLaren, D.G., Callot, V., Cohen-Adad, J., Sdika, M., 2018. Automatic spinal cord localization, robust to MRI contrasts using global curve optimization. Med. Image Anal. 44, 215–227.

Guizard, N., Coupé, P., Fonov, V.S., Manjón, J.V., Arnold, D.L., Collins, D.L., 2015. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. Neuroimage Clin 8, 376–389.

Harmouche, R., Subbanna, N.K., Collins, D.L., Arnold, D.L., Arbel, T., 2015. Probabilistic multiple sclerosis lesion classification based on modeling regional intensity variability and local neighborhood information. IEEE Trans. Biomed. Eng. 62, 1281–1292.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2015. Brain Tumor Segmentation with Deep Neural Networks. arXiv [cs.CV].

Havaei, M., Guizard, N., Chapados, N., Bengio, Y., 2016. HeMIS: Hetero-Modal Image Segmentation, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 469–477.

Healy, B.C., Buckle, G.J., Ali, E.N., Egorova, S., Khalid, F., Tauhid, S., Glanz, B.I., Chitnis, T., Guttmann, C.R.G., Weiner, H.L., Bakshi, R., 2017. Characterizing Clinical and MRI Dissociation in Patients with Multiple Sclerosis. J. Neuroimaging 27, 481–485.

Horsfield, M.A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M.P., Rocca, M.A., Bakshi, R., Filippi, M., 2010. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. Neuroimage 50, 446–455.

Hua, L.H., Donlon, S.L., Sobhanian, M.J., Portner, S.M., Okuda, D.T., 2015. Thoracic spinal cord lesions are influenced by the degree of cervical spine involvement in multiple sclerosis. Spinal Cord 53, 520–525.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790.

Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Kearney, H., Altmann, D.R., Samson, R.S., Yiannakas, M.C., Wheeler-Kingshott, C.A.M., Ciccarelli, O., Miller, D.H., 2015a. Cervical cord lesion load is associated with disability independently from atrophy in MS. Neurology 84, 367–373.

Kearney, H., Miller, D.H., Ciccarelli, O., 2015b. Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. Nat. Rev. Neurol. 11, 327–338.

Kearney, H., Yiannakas, M.C., Abdel-Aziz, K., Wheeler-Kingshott, C.A.M., Altmann, D.R., Ciccarelli, O., Miller, D.H., 2014. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J. Magn. Reson. Imaging 39, 617–623.

Kingma, D., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv [cs.LG].

Koh, J., Kim, T., Chaudhary, V., Dhillon, G., 2010. Automatic segmentation of the spinal cord and the dural sac in lumbar MR images using gradient vector flow field. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2010, 3117–3120.

Krähenbühl, P., Koltun, V., 2011. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, in: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24. Curran Associates, Inc., pp. 109–117.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal Loss for Dense Object Detection. arXiv [cs.CV].

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Little, J.P., Pearcy, M.J., Izatt, M.T., Boom, K., Labrom, R.D., Askin, G.N., Adam, C.J., 2016. Understanding how axial loads on the spine influence segmental biomechanics for idiopathic scoliosis patients: A magnetic resonance imaging study. Clin. Biomech. 32, 220–228.

Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, À., 2012. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. Inf. Sci. 186, 164–185.

Losseff, N.A., Webb, S.L., O'Riordan, J.I., Page, R., Wang, L., Barker, G.J., Tofts, P.S., McDonald, W.I., Miller, D.H., Thompson, A.J., 1996. Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. Brain 119 (Pt 3), 701–708.

Lundell, H., Svolgaard, O., Dogonowski, A.-M., Romme Christensen, J., Selleberg, F., Soelberg Sørensen, P., Blinkenberg, M., Siebner, H.R., Garde, E., 2017. Spinal cord atrophy in anterior-posterior direction reflects impairment in multiple sclerosis. Acta Neurol. Scand. 136, 330–337.

Martin, A.R., De Leener, B., Cohen-Adad, J., Cadotte, D.W., Kalsi-Ryan, S., Lange, S.F., Tetreault, L., Nouri, A., Crawley, A., Mikulis, D.J., Ginsberg, H., Fehlings, M.G., 2017. A Novel MRI Biomarker of Spinal Cord White Matter Injury: T2*-Weighted White Matter to Gray Matter Signal Intensity Ratio. AJNR Am. J. Neuroradiol. https://doi.org/10.3174/ajnr.A5162

Meier, D.S., Guttmann, C.R.G., Tummala, S., Moscufo, N., Cavallari, M., Tauhid, S., Bakshi, R., Weiner, H.L., 2018. Dual-Sensitivity Multiple Sclerosis Lesion and CSF Segmentation for Multichannel 3T Brain MRI. J. Neuroimaging 28, 36–47.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 807–814.

Nakamura, M., Miyazawa, I., Fujihara, K., Nakashima, I., Misu, T., Watanabe, S., Takahashi, T., Itoyama, Y., 2008. Preferential spinal central gray matter involvement in neuromyelitis optica. J. Neurol. 255, 163–170.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42, 1072–1081.

Papinutto, N., Schlaeger, R., Panara, V., Zhu, A.H., Caverzasi, E., Stern, W.A., Hauser, S.L., Henry, R.G., 2015. Age, gender and normalization covariates for spinal cord gray matter and total cross-sectional areas at cervical and thoracic levels: A 2D phase sensitive inversion recovery imaging study. PLoS One 10, e0118576.

Paquin, M.-Ê., El Mendili, M.M., Gros, C., Dupont, S.M., Cohen-Adad, J., Pradat, P.-F., 2018. Spinal Cord Gray Matter Atrophy in Amyotrophic Lateral Sclerosis. AJNR Am. J. Neuroradiol. 39, 184–192.

Perone, C.S., Calabrese, E., Cohen-Adad, J., 2017. Spinal cord gray matter segmentation using deep dilated convolutions. arXiv [cs.CV].

Pezold, S., Fundana, K., Amann, M., Andelova, M., Pfister, A., Sprenger, T., Cattin, P.C., 2015. Automatic Segmentation of the Spinal Cord Using Continuous Max Flow with Cross-sectional Similarity Prior and Tubularity Features, in: Yao, J., Glocker, B., Klinder, T., Li, S. (Eds.), Recent Advances in Computational Methods and Clinical Applications for Spine Imaging, Lecture Notes in Computational Vision and Biomechanics. Springer International Publishing, pp. 107–118.

Popescu, B.F.G., Lucchinetti, C.F., 2012. Pathology of demyelinating diseases. Annu. Rev. Pathol. 7, 185–217.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M., 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Med. Image Comput. Comput. Assist. Interv. 16, 246–253.

Rocca, M.A., Horsfield, M.A., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., Martinelli, V., Caputo, D., Stosic-Opincal, T., Drulovic, J., Comi, G., Filippi, M., 2011. A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. Neurology 76, 2096–2102.

Rocca, M.A., Valsasina, P., Damjanovic, D., Horsfield, M.A., Mesaros, S., Stosic-Opincal, T., Drulovic, J., Filippi, M., 2013. Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. J. Neurol. Neurosurg. Psychiatry 84, 35–41.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, pp. 234–241.

Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T., 2011. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Med. Image Anal. 15, 267–282.

Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2010. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. Neuroimage 49, 1524–1535.

Simard, P.Y., Steinkraus, D., Platt, J.C., Others, 2003. Best practices for convolutional neural networks applied to visual document analysis, in: ICDAR. pp. 958–962.

Simon, J.H., Li, D., Traboulsee, A., Coyle, P.K., Arnold, D.L., Barkhof, F., Frank, J.A., Grossman, R., Paty, D.W., Radue, E.W., Wolinsky, J.S., 2006. Standardized MR imaging protocol for multiple sclerosis: Consortium of MS Centers consensus guidelines. AJNR Am. J. Neuroradiol. 27, 455–461.

Singhal, T., Tauhid, S., Hurwitz, S., Neema, M., Bakshi, R., 2017. The Effect of Glatiramer Acetate on Spinal Cord Volume in Relapsing-Remitting Multiple Sclerosis. J. Neuroimaging 27, 33–36.

Sombekke, M.H., Wattjes, M.P., Balk, L.J., Nielsen, J.M., Vrenken, H., Uitdehaag, B.M.J., Polman, C.H., Barkhof, F., 2013. Spinal cord lesions in patients with clinically isolated syndrome: a powerful tool in diagnosis and prognosis. Neurology 80, 69–75.

Srivastava, N., Hinton, G., Krizhevsky, A., 2014. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine.

Stroman, P.W., Figley, C.R., Cahill, C.M., 2008. Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. Magn. Reson. Imaging 26, 809–814.

Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S., 2008. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. Midas J. 2008, 1–6.

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer International Publishing, pp. 240–248.

Tang, L., Wen, Y., Zhou, Z., von Deneen, K.M., Huang, D., Ma, L., 2013. Reduced field-of-view DTI segmentation of cervical spine tissue. Magn. Reson. Imaging 31, 1507–1514.

Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintoré, M.,

Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C., Cohen, J.A., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 17, 162–173.

Thorpe, J.W., Kidd, D., Moseley, I.F., Kenndall, B.E., Thompson, A.J., MacManus, D.G., McDonald, W.I., Miller, D.H., 1996. Serial gadolinium-enhanced MRI of the brain and spinal cord in early relapsing-remitting multiple sclerosis. Neurology 46, 373–378.

Topfer, R., Foias, A., Rios, N.L., 2018. Integrated ΔB0/Rx coil array for im- proved spinal cord imaging at 3T, in: Proceedings of the 26th Annual Meeting of ISMRM. Presented at the ISMRM, p. 8930.

Topfer, R., Starewicz, P., Lo, K.-M., Metzemaekers, K., Jette, D., Hetherington, H.P., Stikov, N., Cohen-Adad, J., 2016. A 24-channel shim array for the human spinal cord: Design, evaluation, and application. Magn. Reson. Med. 76, 1604–1611.

Ullmann, E., Thong, W., Pelletier Paquette, J.P., Cohen-Adad, J., 2014. Automatic labeling of vertebral levels using a robust template-based approach. Int. J. Biomed. Imaging Article ID 719520, 9 pages.

Valindria, V.V., Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E.O., Rockall, A.G., Rueckert, D., Glocker, B., 2018. Domain Adaptation for MRI Organ Segmentation using Reverse Classification Accuracy.

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Oliver, A., Lladó, X., 2017a. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159–168.

Valverde, S., Oliver, A., Roura, E., González-Villà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Lladó, X., 2017b. Automated tissue segmentation of MR brain images in the presence of white matter lesions. Med. Image Anal. 35, 446–457.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction of MR images of the brain. IEEE Trans. Med. Imaging 18, 885–896.

Vannesjo, S.J., Shi, Y., Tracey, I., Miller, K.L., Clare, S., 2017. Slice-wise first-order shimming of the human spinal cord at 7T, in: Proceedings of the 25th Annual Meeting of ISMRM. p. 5210.

Weier, K., Mazraeh, J., Naegelin, Y., Thoeni, A., Hirsch, J.G., Fabbro, T., Bruni, N., Duyar, H., Bendfeldt, K., Radue, E.-W., Kappos, L., Gass, A., 2012. Biplanar MRI for the assessment of the spinal cord in multiple sclerosis. Mult. Scler. 18, 1560–1569.

Yu, F., Koltun, V., 2015. Multi-Scale Context Aggregation by Dilated Convolutions. arXiv [cs.CV].

Yushkevich, P.A., Gerig, G., 2017. ITK-SNAP: An Intractive Medical Image Segmentation Tool to Meet the Need for Expert-Guided Segmentation of Complex Medical Images. IEEE Pulse 8, 54–57.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S., 2015. Conditional Random Fields as Recurrent Neural Networks, in: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1529–1537.

CHAPTER 5 GENERAL DISCUSSION

In this chapter, we will discuss some limitations of the work done during this Master, and suggest some perspectives and recommendations to further improve it.

5.1 Segmentation

During my Master's project, I was involved in the MRI data management of NeuroPoly Lab. Thanks to international collaborations with clinical or research groups, the number of subjects in the database increased from ~200 to more than 1,500 subjects in two years. This amount of data and its broad variety was essential to train and validate automatic tools, to ensure they generalize well to most of data features.

5.1.1 Spinal cord manual segmentation

Most of these data now have a spinal cord manual segmentation and MS patients lesion manual segmentation. To do so, we mainly used FSLview (Jenkinson et al., 2012) and ITK-Snap (Yushkevich and Gerig, 2017) software, the latest having the advantage of a useful intensity-based tool which propagate a segmentation from a seed set by a user. It should be noted the semi-automatic nature of the manual segmentation since we have manually corrected the automatic segmentation output by *PropSeg* (De Leener et al., 2014) and considered it as "ground truth". It certainly introduced a bias which should be quantified by comparing then intra-rater variability of the fully-manual spinal cord segmentation to semi-manual segmentation (i.e when using *PropSeg* initialization).

5.1.2 Intramedullary MS lesions manual segmentation

MS lesion were segmented by 7 raters including radiologists and trained raters. I would like to take the opportunity to warmly thank the raters for their tremendous contribution, this project would not have been possible without their rigour and expertise.

Guidelines were provided to the raters so that reduce the inter-variability (https://osf.io/d4evy/). For instance, when both 2D axial and sagittal scans were available for a subject, raters were urged to start segmenting the sagittal scan, then go to the axial scan, and if necessary, cross-check identified lesions across views. Having two images per subjects helped them to reduce the number of false positive detections by taking the best of both worlds: the good detectability of lesions in the sagittal plane of the sagittal images as well as the good contrast of

lesion in the cross-sectional plane of axial images. For this project, we asked the raters to minimize false positives (i.e. only segment voxels for which we are almost certain they belong to as lesion). The same applies for lesion edge: better to not label a voxel if we are not sure it is a lesion. However, important efforts have to be made to reduce the variability across raters (see **Figure 4.7**). To this end, it would be important to establish more extensive consensual segmentation guidelines for the intramedullary MS lesion segmentation, in order to standardize the validation of automatic segmentation methods. For instance, it could be valuable to set a minimum lesion size (e.g. 3 voxels) to label a lesion since the clinical significance of little dots of lesion is unclear. Manual lesion segmentation protocols established for brain lesions could be of interest (e.g. [URL]).

5.1.3 Validation metrics

For the inter-rater study, I used the majority voting method to compute the consensus masks across raters. Further studies should consider instead the simultaneous truth and performance level estimation (STAPLE) algorithm (Warfield et al. 2004), as it was done recently for the ISBI 2015 brain lesion segmentation challenge (Carass et al., 2017). Briefly, STAPLE provides a probabilistic estimate of the "true" segmentation from an optimal combination (computed by the expectation maximization algorithm) of the input masks.

The quality of a segmentation can be evaluated in many different ways, but I believe there is a need for establishing a consensus about the validation metrics to assess the ability of an automatic method to segment intramedullary lesions. Again, we should take advantage of the work that has been performed for brain lesions (Styner et al. 2008; Carass et al. 2017). Besides the metrics used in **Chapter 4**, the following metrics could be considered: (i) positive predictive value (voxel-wise ratio of the true positives to the sum of true and false positives), (ii) lesion-wise specificity (which is important for the medical community in the context of clinical trials), (iii) the average symmetric surface distance, (iv) volume correlation (Pearson's correlation coefficient between the manual and automatic volumes. Similarly, for the spinal cord segmentation, additional metrics should be of interest, such as distance metrics between the contours to better evaluate under/over segmentation cases.

5.1.4 Open-source multi-center database

The spinal cord research community could benefit from an open-source multi-center database, as it was the case under the initiative of the "*Spinal cord grey matter segmentation challenge*" (Prados et al., 2017) or even more largely with the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008). The motivation behind this would be to create a consensus of the validation process of automatic segmentation methods applied to the spinal cord and intramedullary lesions, as it was done for the grey matter with a complete validation framework (Prados et al., 2017) or for brain MS lesions (Styner et al., 2008). Similar initiatives could also address the question of acquisition standardization to reduce variations in sequences and image contrast scanning spinal cord of MS patients, as it was recently proposed for healthy subjects by Alley et al. (Alley et al., 2018).

5.2 Post-processing

The framework presented in this thesis does not contain additional post-processing. Post-processing steps are sometimes added to refine the segmentation output by the CNNs. Graphical models, such as Markov Random Fields (MRF) or Conditional Random Fields (CRF), were recently applied on top of the probabilistic map provided by the CNNs, to regularized the labels (Alansary et al., 2016; Cai et al., 2016; Christ et al., 2016; Gao et al., 2016; Krähenbühl and Koltun, 2011; Shakeri et al., 2016; Song et al., 2015). An example of their utilisation as post-processing is illustrated in **Figure 5.1 2**. The motivation behind the use of graphical models for medical image segmentation is that anatomical structures are organized as continuous and coherent shapes, and then sites strongly depend on their neighborhood.

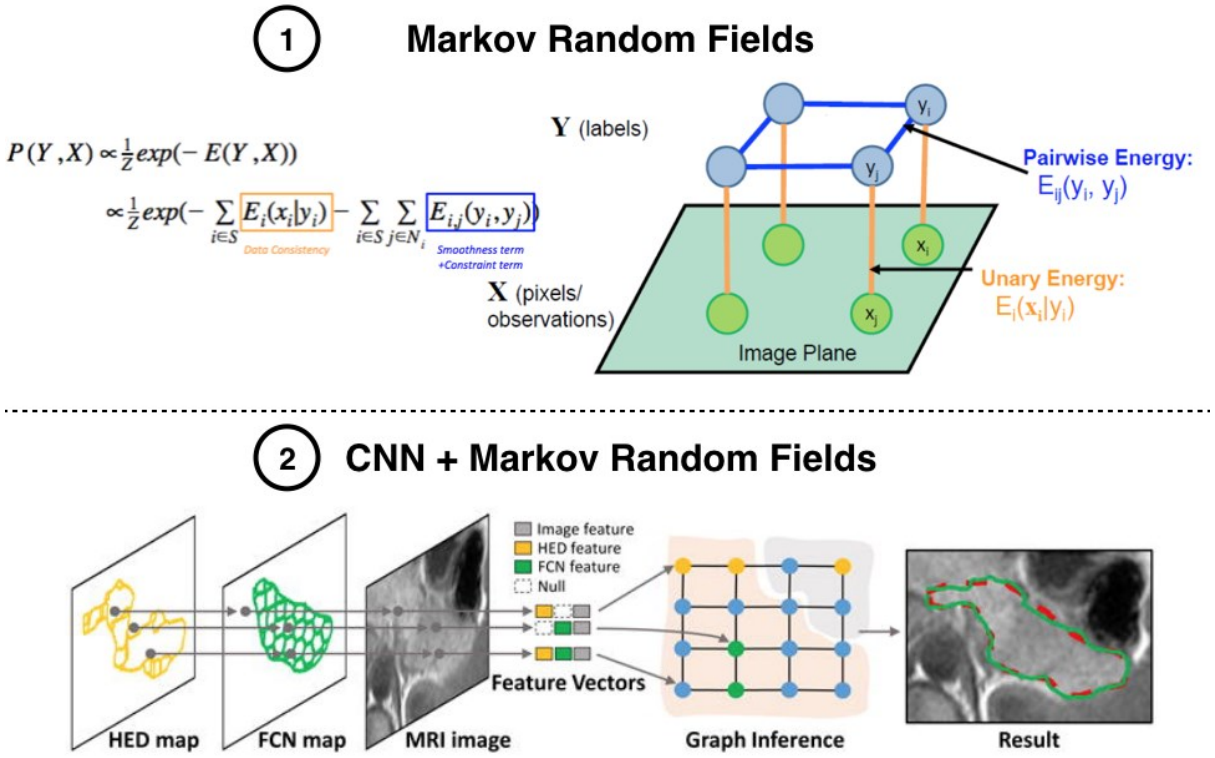


Figure 5.1: Markov Random Fields.

(1) shows a diagram representation of Markov Random Fields as graphical model. Each node represents either a data feature (in green, e.g. pixel intensity) or a label (in blue, i.e. predicted label of each image pixel), while the edges represents their dependencies. The joint probability (left) depends on the unary (dependence between a pixel and its label) and the pairwise (relationships between labels of neighboring sites) energies. The unary term ensures the data consistency while the pairwise energy can be considered as a spatial regularization term (i.e. encourage similar labels between neighboring pixels). (2) illustrates an utilization of graphical models as post-processing after CNN prediction. Here, the information of two CNN predictions (HED and FCN maps) and raw image are combined in the graph nodes, which are then partitioned by graph inference. (1) is adapted from Zahra Karimaghaloo slides, (2) is extracted from (Cai et al., 2016).

Briefly, MRF or CRF can be used for image segmentation as probabilistic graphical model. Local contextual information is encoded thanks to the definition of a neighbourhood system and potential functions (Li, 2009). As shown in **Figure 5.1 1**, considering the labels and the image as random fields, we want to infer the most probable label configuration given the

observed data. As a result, MRF regularization adds powerful spatial regularization to independent and identically distributed classification. The problem is equivalent to an energy minimization problem, of which the exact computation of the most probable configuration is intractable. We must resort to approximation methods to solve it (Bauer et al., 2011; Karimaghaloo et al., 2012; Van Leemput et al., 2001). For instance, Christ et al. formulated the final label assignment given the probability maps of the CNN as maximum a posteriori inference in a dense CRF (Christ et al., 2016), as previously suggested by (Krähenbühl and Koltun, 2011).

These graphical models come with additional hyperparameters, which are required to be fine-tuned (very time-consuming and task dependent). Moreover, finding a set of hyper-parameters that works for every dataset is challenging. A recent study proposed a solution to overcome this issue by successfully integrating the learning of the CRF hyperparameters within the training pipeline (Zheng et al., 2015)

5.3 Multi-stream approach

In clinical routine, it is not uncommon to have more than one spinal cord acquisition. In the proposed solution of this Master's project, MS lesion segmentation was however achieved by processing each 3D scan independently. This is arguably a sub-optimal use of the different images available per patients since a combination of the information from different MR contrasts should help the identification of very small lesions while reducing the number of false positives. In particular, Kearney et al. states that a combination of both sagittal and axial MRI sequences could largely improve the identification of focal lesions and diffuse abnormalities (Kearney et al., 2015b). A solution would be to register both the sagittal and axial images in a common normalized space, such as the PAM50 template (De Leener et al., 2018) and then to perform the segmentation considering each image as a different input channel. Register the MRI scans is actually the typical approach used to segment MS lesions in the brain (García-Lorenzo et al., 2013; Lladó et al., 2012), and is also used by the very recent solutions using CNN combining T_1-w, T_2-w and FLAIR images (Roy et al., 2018; Valverde et al., 2017). It should be noted that, unlike brain images, spinal cord acquisitions of one patient usually do not cover the same field of view (e.g. the Superior-to-Inferior coverage of axial scans is usually much smaller than sagittal scans), and present substantial differences in voxel sizes, which could hinder a consistent registration between images.

In multi-center studies, it is however not unusual to observe an heterogeneity in terms of available MR sequences (e.g. center #1: {T2*-w and T2-w}, center #2: {T1-w}, center #3: {T2-w}). Hence, a solution combining the information of different sequence needs to be robust to missing sequences. Havaei et al. proposed a segmentation framework which is robust to incomplete multi-modal datasets, as illustrated in **Figure 5.2** (Havaei et al., 2016). In their approach, the available scans are combined in an "Abstraction" space by arithmetic operations (such as mean operations) on feature maps resulting from prior specific-modality convolutional layers. Thus, they devised a framework able to combine multi-modal information while being robust to any combinatorial subset of available input modalities.

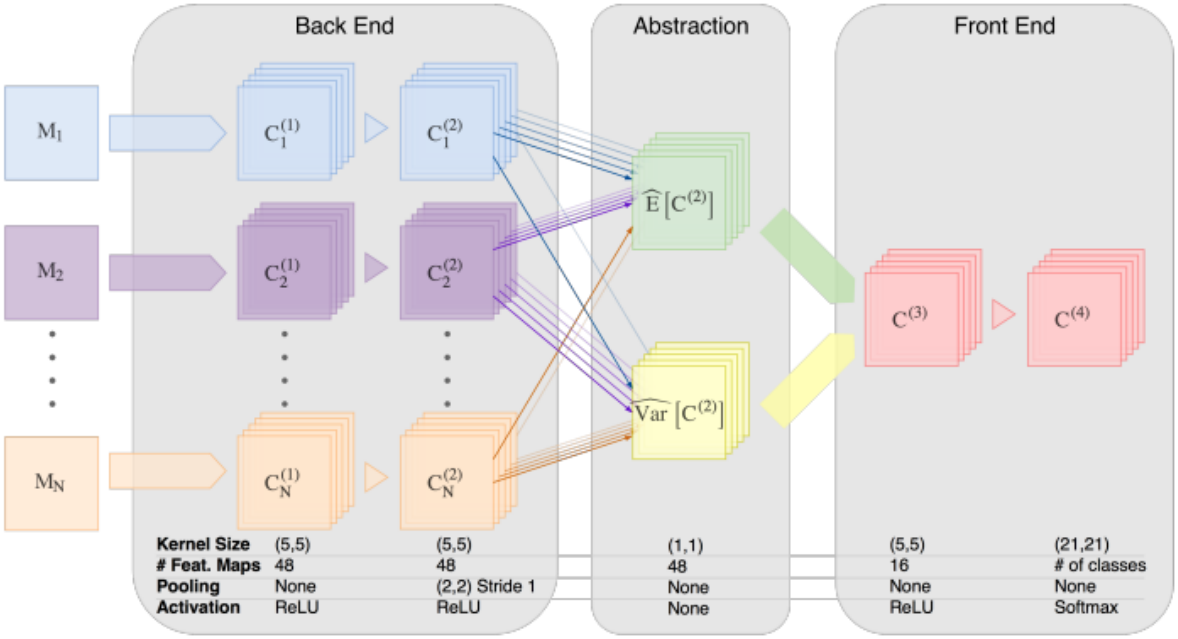


Figure 5.2: Multi-stream approach.

Illustration of the segmentation framework proposed by Havaei et al., which is robust to missing modalities. In the "back end" (left), each modality, M_k, is processed independently by modality-specific convolutional layers. In the "abstraction" (middle) part, feature maps statistics are computed from the "back end" feature maps. These maps are then all concatenated in the "front end" part, where final convolutional layers provide the output segmentation. Figure extracted from (Havaei et al., 2016).

5.4 CNN training

5.4.1 General considerations

After months tweaking CNN trainings on Jupyter Notebook, after dozens and dozens of false expectations when staring at TensorBoard slopes, I wished to share my personal feeling/feedback, hoping it could help somehow a curious reader:

- The exact architecture is not the most important aspect of the training. However, I believe of the great potential of designing a framework specifically devised for a task, combined with a good knowledge about the anatomical structure to segment and the variety of the data features.
- An appropriated data pre-processing and/or augmentation could lead to significant improvements of the results and to networks more robust.
- Concerning the model hyperparameter optimization, I have the regret (or the satisfaction) to say that there is no clear recipe. To obtain satisfactory hyper-parameter set is time-consuming and empirical task, even if some rules of thumb / intuition-based random search exists (Bengio, 2012; Bergstra and Bengio, 2012).
- I feel extremely lucky to do my research in an area where "code in open-access" is a watchword. Blogs, workshops, debates, new developments, code readily available and each day enriched by an active community is an extraordinary opportunity to learn. Let's try to contribute to it as well…

5.4.2 Domain adaptation

A main challenge of this project was the substantial variations in imaging protocols. For instance, MRI does not have any uniform intensity scale (as CT imaging). To face this issue, recent applications of domain-adaptation in medical image area could be of interest. Domain adaptation is a special case of transfer learning, which is the process of improving the learning of the segmentation in the target domain (e.g. MRI data of center #3), using the information learned from labeled data in one or more source domains (e.g. center #1 and #2) (Pan and Yang, 2010). An application of this concept with supervised CNNs has been proposed by Ghafoorian et al. (Ghafoorian et al., 2017), and illustrated on **Figure 5.3**.

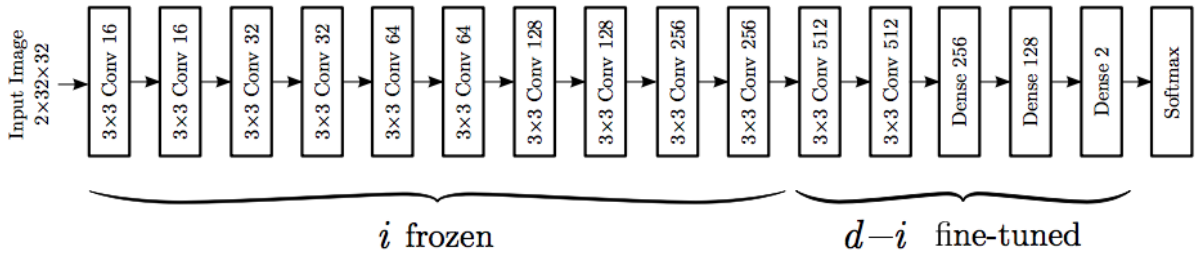


Figure 5.3: Domain Adaptation.

A model, with d layers, is first trained on data of the source domain. The illustrated network is a predictive model initially trained on the source domain (i.e. with transfer of learned weights), and domain-adapted to the target domain by (i) frozing the i first layers and (ii) fine-tuning the remaining d-i last layers with the data from the target domain. Figure extracted from (Ghafoorian et al., 2017).

5.4.3 Semi-supervised methods

Despite the expensive cost of data acquisition in medical imaging domain, Litjens et al. reported that, sometimes, supervised methods are limited by the lack of labeled images while unlabeled data are available (Litjens et al., 2017b). Indeed, image labelling is time consuming and tedious, but is essential for the training of supervised methods. A large amount of labeled data is especially crucial for supervised CNN methods to prevent overfitting. Semi-supervised or unsupervised methods could be of interest for taking advantage of unlabeled data. Semi-supervised methods are trained on dataset with some labeled data, and the remaining without available labeling. The ambition of semi-supervised methods lies in: if the unlabeled data could provide useful problem for the inference problem, then we can expect semi-supervised learning to yield improvements compared to typical supervised learning methods (Baur et al., 2017; Jalilian et al., 2017; Weston et al., 2012).

CHAPTER 6 CONCLUSION AND RECOMMENDATIONS

Over the past three years, I have been part of the research activity of the NeuroPoly Lab, at Ecole Polytechnique of Montreal, under the supervision of Prof. Julien Cohen-Adad. My work has gravitated around the image processing and analysis of spinal cord structural MRI data, which mainly resulted in a fully automated method to segment the intramedullary MS lesions. This framework is made of a sequence of two convolutional neural networks and has been validated against a multi-center, -contrast, and -resolution data set. To encourage reproducibility, further developments, and utilization of the proposed method, its implementation is integrated in the SCT open-source software.

My recommendations to further improve the proposed method would be:

1. To increase the interactions with neuro-radiologists:
    a. To establish more rigorous and consensual guidelines dedicated to the manual segmentation of intramedullary MS lesions, in order to reduce the inter- and intra-rater variability.
    b. To integrate more discriminative clinical and anatomical knowledge in the automatic segmentation method.
2. To extend the automatic detection to other pathologies than MS.
3. To estimate the uncertainty of the produced segmentation to allow subsequent revision by clinicians.
4. To implement a training process to permit other research groups to fine-tune the segmentation model for their particular data sets.
5. To allow the framework to combine multiple MR scans of a same patient.

REFERENCES

Ahuja, C.S., Wilson, J.R., Nori, S., Kotter, M.R.N., Druschel, C., Curt, A., Fehlings, M.G., 2017. Traumatic spinal cord injury. Nat Rev Dis Primers 3, 17018.

Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., Rutherford, M., Hajnal, J.V., Glocker, B., Rueckert, D., Kainz, B., 2016. Fast Fully Automatic Segmentation of the Human Placenta from Motion Corrupted MRI, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 589–597.

Alley, S., Gilbert, G., Gandini Wheeler-Kingshott, C.A., Samson, R.S., Grussu, F., Martin, A.R., Bannier, E., Callot, V., Smith, S.A., Xu, S., Dewey, B., Weber, K.A., Parrish, T.B., McLaren, D., Barker, G.J., Papinutto, N., Seif, M., Freund, P., Barry, R.L., By, S., Narayanan, S., Cohen-Adad, J., 2018. Consensus acquisition protocol for quantitative MRI of the cervical spinal cord at 3T, in: Proceedings of the 26th Annual Meeting of ISMRM. Presented at the ISMRM.

Arrambide, G., Rovira, A., Sastre-Garriga, J., Tur, C., Castilló, J., Río, J., Vidal-Jordana, A., Galán, I., Rodríguez-Acevedo, B., Midaglia, L., Nos, C., Mulero, P., Arévalo, M.J., Comabella, M., Huerga, E., Auger, C., Montalban, X., Tintore, M., 2018. Spinal cord lesions: A modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor. Mult. Scler. 24, 301–312.

Bakshi, R., Dandamudi, V.S.R., Neema, M., De, C., Bermel, R.A., 2005. Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. J. Neuroimaging 15, 30S–45S.

Bastianello, S., Paolillo, A., Giugni, E., Giuliani, S., Evangelisti, G., Luccichenti, G., Angeloni, U., Colonnese, C., Salvetti, M., Gasperini, C., Pozzilli, C., Fieschi, C., 2000. MRI of spinal cord in MS. J. Neurovirol. 6 Suppl 2, S130–3.

Bauer, S., Nolte, L.-P., Reyes, M., 2011. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. Med. Image Comput. Comput. Assist. Interv. 14, 354–361.

Baur, C., Albarqouni, S., Navab, N., 2017. Semi-supervised Deep Learning for Fully Convolutional Networks, in: Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. Springer International Publishing, pp. 311–319.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008/6. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110, 346–359.

Bengio, Y., 2012. Practical Recommendations for Gradient-Based Training of Deep Architectures, in: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade: Second Edition. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 437–478.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Bhole, C., Morsillo, N., Pal, C., 2011. 3D Segmentation in CT Imagery with Conditional Random Fields and Histograms of Oriented Gradients, in: Suzuki, K., Wang, F., Shen, D., Yan, P. (Eds.), Machine Learning in Medical Imaging, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 326–334.

Biberacher, V., Boucard, C.C., Schmidt, P., Engl, C., Buck, D., Berthele, A., Hoshi, M.-M., Zimmer, C., Hemmer, B., Mühlau, M., 2015. Atrophy and structural variability of the upper cervical cord in early multiple sclerosis. Mult. Scler. 21, 875–884.

Blumenfeld, H., n.d. Neuroanatomy through clinical cases. ; 2010. Sunderland, MA: Sinauer Associates.

Bot, J.C.J., Blezer, E.L.A., Kamphorst, W., Lycklama A Nijeholt, G.J., Ader, H.J., Castelijns, J.A., Ig, K.N., Bergers, E., Ravid, R., Polman, C., Barkhof, F., 2004. The spinal cord in multiple sclerosis: relationship of high-spatial-resolution quantitative MR imaging findings to histopathologic results. Radiology 233, 531–540.

Breckwoldt, M.O., Gradl, J., Hähnel, S., Hielscher, T., Wildemann, B., Diem, R., Platten, M., Wick, W., Heiland, S., Bendszus, M., 2017. Increasing the sensitivity of MRI for the detection of multiple sclerosis lesions by long axial coverage of the spinal cord: a prospective study in 119 patients. J. Neurol. 264, 341–349.

Brooks, B.R., 1996. Natural history of ALS: symptoms, strength, pulmonary function, and disability. Neurology 47, S71–81; discussion S81–2.

Brooks, B.R., 1994. El escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. J. Neurol. Sci. 124, 96–107.

Brosch, T., Tang, L.Y.W., Youngjin Yoo, Li, D.K.B., Traboulsee, A., Tam, R., 2016. Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. IEEE Trans. Med. Imaging 35, 1229–1239.

Buades, A., Coll, B., Morel, J., 2005. A Review of Image Denoising Algorithms, with a New One. Multiscale Model. Simul. 4, 490–530.

Buda, M., Maki, A., Mazurowski, M.A., 2017. A systematic study of the class imbalance problem in convolutional neural networks. arXiv preprint arXiv:1710. 05381.

Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q., 2016. Pancreas Segmentation in MRI using Graph-Based Decision Fusion on Convolutional Neural Networks. Med. Image Comput. Comput. Assist. Interv. 9901, 442–450.

Carbonell-Caballero, J., Manjón, J.V., Martí-Bonmatí, L., Olalla, J.R., Casanova, B., de la Iglesia-Vayá, M., Coret, F., Robles, M., 2006. Accurate quantification methods to evaluate cervical cord atrophy in multiple sclerosis patients. MAGMA 19, 237–246.

Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A.M., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J. Magn. Reson. Imaging 32, 223–228.

Chen, M., Carass, A., Oh, J., Nair, G., Pham, D.L., Reich, D.S., Prince, J.L., 2013. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. Neuroimage 83, 1051–1062.

Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M., Sommer, W.H., Ahmadi, S.-A., Menze, B.H., 2016. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 415–423.

Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Bruce Pike, G., Evans, A.C., 1997. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database, in: NeuroImage.

Cohen-Adad, J., El Mendili, M.-M., Morizot-Koutlidis, R., Lehéricy, S., Meininger, V., Blancho, S., Rossignol, S., Benali, H., Pradat, P.-F., 2013. Involvement of spinal sensory pathway in ALS and specificity of cord atrophy to lower motor neuron degeneration. Amyotroph. Lateral Scler. Frontotemporal Degener. 14, 30–38.

Compston, A., Coles, A., 2002. Multiple sclerosis. Lancet 359, 1221–1231.

Coupé, P., Yger, P., Barillot, C., 2006. Fast non local means denoising for 3D MR images. Med. Image Comput. Comput. Assist. Interv. 9, 33–40.

Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans. Med. Imaging 27, 425–441.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). pp. 886–893 vol. 1.

De Leener, B., Cohen-Adad, J., Kadoury, S., 2015. Automatic Segmentation of the Spinal Cord and Spinal Canal Coupled With Vertebral Labeling. IEEE Trans. Med. Imaging 34, 1705–1718.

De Leener, B., Fonov, V.S., Collins, D.L., Callot, V., Stikov, N., Cohen-Adad, J., 2018. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. Neuroimage 165, 170–179.

De Leener, B., Kadoury, S., Cohen-Adad, J., 2014. Robust, accurate and fast automatic segmentation of the spinal cord. Neuroimage 98, 528–536.

De Leener, B., Lévy, S., Dupont, S.M., Fonov, V.S., Stikov, N., Louis Collins, D., Callot, V., Cohen-Adad, J., 2017. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 145, 24–43.

De Leener, B., Taso, M., Cohen-Adad, J., Callot, V., 2016. Segmentation of the human spinal cord. MAGMA 29, 125–153.

Diaz, I., Boulanger, P., Greiner, R., Murtha, A., 2011. A critical review of the effects of de-noising algorithms on MRI brain tumor segmentation. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2011, 3934–3937.

Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. Ecology 26, 297–302.

Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T., 2014. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp. 766–774.

Dupont, S.M., De Leener, B., Taso, M., Le Troter, A., Nadeau, S., Stikov, N., Callot, V., Cohen-Adad, J., 2017. Fully-integrated framework for the segmentation and registration of the spinal cord white and gray matter. Neuroimage 150, 358–372.

El Mendili, M.-M., Cohen-Adad, J., Pelegrini-Issac, M., Rossignol, S., Morizot-Koutlidis, R., Marchand-Pauvert, V., Iglesias, C., Sangari, S., Katz, R., Lehericy, S., Benali, H., Pradat, P.-F., 2014. Multi-parametric spinal cord MRI as potential progression marker in amyotrophic lateral sclerosis. PLoS One 9, e95516.

Fog, T., 1950. Topographic distribution of plaques in the spinal cord in multiple sclerosis. Archives of Neurology & Psychiatry 63, 382–414.

Fradet, L., Arnoux, P.-J., Ranjeva, J.-P., Petit, Y., Callot, V., 2014. Morphometrics of the entire human spinal cord and spinal canal measured from in vivo high-resolution anatomical magnetic resonance imaging. Spine 39, E262–9.

Freund, P., Weiskopf, N., Ashburner, J., Wolf, K., Sutter, R., Altmann, D.R., Friston, K., Thompson, A., Curt, A., 2013. MRI investigation of the sensorimotor cortex and the corticospinal tract after acute spinal cord injury: a prospective longitudinal study. Lancet Neurol. 12, 873–881.

Gao, M., Xu, Z., Lu, L., Wu, A., Nogues, I., Summers, R.M., Mollura, D.J., 2016. Segmentation label propagation using deep convolutional neural networks and dense conditional random field, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, pp. 1265–1268.

García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L., 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med. Image Anal. 17, 1–18.

Gelineau-Morel, R., Tomassini, V., Jenkinson, M., Johansen-Berg, H., Matthews, P.M., Palace, J., 2012. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum. Brain Mapp. 33, 2802–2814.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.-E., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., Wells, W.M., 2017. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation, in: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 516–524.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goto, N., Otsuka, N., 1997. Development and anatomy of the spinal cord. Neuropathology 17, 25–31.

Gros, C., De Leener, B., Dupont, S.M., Martin, A.R., Fehlings, M.G., Bakshi, R., Tummala, S., Auclair, V., McLaren, D.G., Callot, V., Cohen-Adad, J., Sdika, M., 2018. Automatic spinal cord localization, robust to MRI contrasts using global curve optimization. Med. Image Anal. 44, 215–227.

Gros, C., De Leener, B., Dupont, S.M., Martin, A.R., Fehlings, M.G., Bakshi, R., Tummala, S., Auclair, V., McLaren, D.G., Callot, V., Sdika, M., Cohen-Adad, J., 2017. OptiC: Robust and Automatic Spinal Cord Localization on a Large Variety of MRI Data Using a Distance Transform Based Global

Optimization, in: Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, Lecture Notes in Computer Science. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, pp. 712–719.

Harmouche, R., Subbanna, N.K., Collins, D.L., Arnold, D.L., Arbel, T., 2015. Probabilistic multiple sclerosis lesion classification based on modeling regional intensity variability and local neighborhood information. IEEE Trans. Biomed. Eng. 62, 1281–1292.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2015. Brain Tumor Segmentation with Deep Neural Networks. arXiv [cs.CV].

Havaei, M., Guizard, N., Chapados, N., Bengio, Y., 2016. HeMIS: Hetero-Modal Image Segmentation, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 469–477.

Herskovits, n.d. Accuracy for detection of simulated lesions: comparison of fluid-attenuated inversion-recovery, proton density-weighted, and T2-weighted synthetic brain MR imaging.

He, X., Zemel, R.S., Carreira-Perpiñán, M.Á., 2004. Multiscale conditional random fields for image labeling, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, pp. II–II.

Hori, M., Hagiwara, A., Fukunaga, I., Ueda, R., Kamiya, K., Suzuki, Y., Liu, W., Murata, K., Takamura, T., Hamasaki, N., Irie, R., Kamagata, K., Kumamaru, K.K., Suzuki, M., Aoki, S., 2018. Application of Quantitative Microstructural MR Imaging with Atlas-based Analysis for the Spinal Cord in Cervical Spondylotic Myelopathy. Sci. Rep. 8, 5213.

Horsfield, M.A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M.P., Rocca, M.A., Bakshi, R., Filippi, M., 2010. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. Neuroimage 50, 446–455.

Hua, L.H., Donlon, S.L., Sobhanian, M.J., Portner, S.M., Okuda, D.T., 2015. Thoracic spinal cord lesions are influenced by the degree of cervical spine involvement in multiple sclerosis. Spinal Cord 53, 520–525.

Ingle, G.T., Stevenson, V.L., Miller, D.H., Thompson, A.J., 2003. Primary progressive multiple sclerosis: a 5-year clinical and MR study. Brain 126, 2528–2536.

Jack, C.R., Jr, Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J.,

Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Jalilian, E., Uhl, A., Kwitt, R., 2017. Domain Adaptation for CNN Based Iris Segmentation. BIOSIG 2017.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790.

Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Karimaghaloo, Z., Shah, M., Francis, S.J., Arnold, D.L., Collins, D.L., Arbel, T., 2012. Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. IEEE Trans. Med. Imaging 31, 1181–1194.

Kearney, H., Altmann, D.R., Samson, R.S., Yiannakas, M.C., Wheeler-Kingshott, C.A.M., Ciccarelli, O., Miller, D.H., 2015a. Cervical cord lesion load is associated with disability independently from atrophy in MS. Neurology 84, 367–373.

Kearney, H., Miller, D.H., Ciccarelli, O., 2015b. Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. Nat. Rev. Neurol. 11, 327–338.

Kearney, H., Miszkiel, K.A., Yiannakas, M.C., Ciccarelli, O., Miller, D.H., 2013. A pilot MRI study of white and grey matter involvement by multiple sclerosis spinal cord lesions. Mult. Scler. Relat. Disord. 2, 103–108.

Kearney, H., Yiannakas, M.C., Abdel-Aziz, K., Wheeler-Kingshott, C.A.M., Altmann, D.R., Ciccarelli, O., Miller, D.H., 2014. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J. Magn. Reson. Imaging 39, 617–623.

Kilsdonk, I.D., Jonkman, L.E., Klaver, R., van Veluw, S.J., Zwanenburg, J.J.M., Kuijer, J.P.A., Pouwels, P.J.W., Twisk, J.W.R., Wattjes, M.P., Luijten, P.R., Barkhof, F., Geurts, J.J.G., 2016. Increased cortical grey matter lesion detection in multiple sclerosis with 7 T MRI: a post-mortem verification study. Brain 139, 1472–1481.

Kingma, D., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv [cs.LG].

Koch, L.M., Rajchl, M., Bai, W., Baumgartner, C.F., Tong, T., Passerat-Palmbach, J., Aljabar, P., Rueckert, D., 2017. Multi-Atlas Segmentation using Partially Annotated Data: Methods and Annotation Strategies. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.2017.2711020

Koh, J., Kim, T., Chaudhary, V., Dhillon, G., 2010. Automatic segmentation of the spinal cord and the dural sac in lumbar MR images using gradient vector flow field. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2010, 3117–3120.

Koh, J., Scott, P.D., Chaudhary, V., Dhillon, G., 2011. An automatic segmentation method of the spinal canal from clinical MR images based on an attention model and an active contour model, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1467–1471.

Krähenbühl, P., Koltun, V., 2011. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, in: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24. Curran Associates, Inc., pp. 109–117.

Lévy, S., Benhamou, M., Naaman, C., Rainville, P., Callot, V., Cohen-Adad, J., 2015. White matter atlas of the human spinal cord with estimation of partial volume effect. Neuroimage 119, 262–271.

Li, S.Z., 2009. Markov Random Field Modeling in Image Analysis. Springer Science & Business Media.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017a. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017b. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, À., 2012. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. Inf. Sci. 186, 164–185.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Losseff, N.A., Webb, S.L., O'Riordan, J.I., Page, R., Wang, L., Barker, G.J., Tofts, P.S., McDonald, W.I., Miller, D.H., Thompson, A.J., 1996. Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. Brain 119 (Pt 3), 701–708.

Lovas, G., Szilágyi, N., Majtényi, K., Palkovits, M., Komoly, S., 2000. Axonal changes in chronic demyelinated cervical spinal cord plaques. Brain 123 (Pt 2), 308–317.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. 60, 91–110.

Lundell, H., Svolgaard, O., Dogonowski, A.-M., Romme Christensen, J., Selleberg, F., Soelberg Sørensen, P., Blinkenberg, M., Siebner, H.R., Garde, E., 2017. Spinal cord atrophy in anterior-posterior direction reflects impairment in multiple sclerosis. Acta Neurol. Scand. 136, 330–337.

Lycklama, G., Thompson, A., Filippi, M., Miller, D., Polman, C., Fazekas, F., Barkhof, F., 2003. Spinal-cord MRI in multiple sclerosis. Lancet Neurol. 2, 555–562.

Maleki, M., Teshnehlab, M., Nabavi, M., 2012. Diagnosis of Multiple Sclerosis (MS) Using Convolutional Neural Network (CNN) from MRIs. Global Journal of Medicinal Plant Research 1, 50–54.

Martin, A.R., De Leener, B., Cohen-Adad, J., Cadotte, D.W., Kalsi-Ryan, S., Lange, S.F., Tetreault, L., Nouri, A., Crawley, A., Mikulis, D.J., Ginsberg, H., Fehlings, M.G., 2017. A Novel MRI Biomarker of Spinal Cord White Matter Injury: T2*-Weighted White Matter to Gray Matter Signal Intensity Ratio. AJNR Am. J. Neuroradiol. https://doi.org/10.3174/ajnr.A5162

Matsuzaki, H., Wakabayashi, K., Ishihara, K., Ishikawa, H., Kawabata, H., Onomura, T., 1996. The origin and significance of spinal cord pulsation. Spinal Cord 34, 422–426.

Melhem, n.d. Defining Thresholds for Changes in Size of Simulated T2-Hyperintense Brain Lesions on the Basis of Qualitative Comparisons.pdf.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571.

MS Society of Canada [WWW Document], n.d. URL https://mssociety.ca/about-ms (accessed 16).

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 807–814.

Nakamura, M., Miyazawa, I., Fujihara, K., Nakashima, I., Misu, T., Watanabe, S., Takahashi, T., Itoyama, Y., 2008. Preferential spinal central gray matter involvement in neuromyelitis optica. J. Neurol. 255, 163–170.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42, 1072–1081.

Nyúl, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale standardization. IEEE Trans. Med. Imaging 19, 143–150.

Oliveira, A., Pereira, S., Silva, C.A., 2017. Augmenting data when training a CNN for retinal vessel segmentation: How to warp?, in: Bioengineering (ENBENG), 2017 IEEE 5th Portuguese Meeting on. IEEE, pp. 1–4.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359.

Papinutto, N., Schlaeger, R., Panara, V., Zhu, A.H., Caverzasi, E., Stern, W.A., Hauser, S.L., Henry, R.G., 2015. Age, gender and normalization covariates for spinal cord gray matter and total cross-sectional areas at cervical and thoracic levels: A 2D phase sensitive inversion recovery imaging study. PLoS One 10, e0118576.

Paquin, M.-Ê., El Mendili, M.M., Gros, C., Dupont, S.M., Cohen-Adad, J., Pradat, P.-F., 2018. Spinal Cord Gray Matter Atrophy in Amyotrophic Lateral Sclerosis. AJNR Am. J. Neuroradiol. 39, 184–192.

Pereira, S., Pinto, A., Alves, V., Silva, C.A., 2016. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans. Med. Imaging 35, 1240–1251.

Perone, C.S., Calabrese, E., Cohen-Adad, J., 2017. Spinal cord gray matter segmentation using deep dilated convolutions. arXiv [cs.CV].

Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69, 292–302.

Popescu, B.F.G., Lucchinetti, C.F., 2012. Pathology of demyelinating diseases. Annu. Rev. Pathol. 7, 185–217.

Prados, F., Ashburner, J., Blaiotta, C., Brosch, T., Carballido-Gamio, J., Cardoso, M.J., Conrad, B.N., Datta, E., Dávid, G., Leener, B.D., Dupont, S.M., Freund, P., Wheeler-Kingshott, C.A.M.G., Grussu, F., Henry, R., Landman, B.A., Ljungberg, E., Lyttle, B., Ourselin, S., Papinutto, N., Saporito, S., Schlaeger, R., Smith, S.A., Summers, P., Tam, R., Yiannakas, M.C., Zhu, A., Cohen-Adad, J., 2017. Spinal cord grey matter segmentation challenge. Neuroimage 152, 312–329.

Renton, A.E., Chiò, A., Traynor, B.J., 2014. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17, 17–23.

Rocca, M.A., Horsfield, M.A., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., Martinelli, V., Caputo, D., Stosic-Opincal, T., Drulovic, J., Comi, G., Filippi, M., 2011. A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. Neurology 76, 2096–2102.

Rocca, M.A., Valsasina, P., Damjanovic, D., Horsfield, M.A., Mesaros, S., Stosic-Opincal, T., Drulovic, J., Filippi, M., 2013. Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. J. Neurol. Neurosurg. Psychiatry 84, 35–41.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, pp. 234–241.

Rossignol, S., 2006. Plasticity of connections underlying locomotor recovery after central and/or peripheral lesions in the adult mammals. Philos. Trans. R. Soc. Lond. B Biol. Sci. 361, 1647–1671.

Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Lladó, X., 2015. A toolbox for multiple sclerosis lesion segmentation. Neuroradiology 57, 1031–1043.

Roy, S., Butman, J.A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2018. Multiple Sclerosis Lesion Segmentation from Brain MRI via Fully Convolutional Neural Networks. arXiv [cs.CV].

Roy, S., Carass, A., Prince, J.L., 2013. PATCH BASED INTENSITY NORMALIZATION OF BRAIN MR IMAGES. Proc. IEEE Int. Symp. Biomed. Imaging 2013, 342–345.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533.

Schlaeger, R., Papinutto, N., Panara, V., Bevan, C., Lobach, I.V., Bucci, M., Caverzasi, E., Gelfand, J.M., Green, A.J., Jordan, K.M., Stern, W.A., von Büdingen, H.-C., Waubant, E., Zhu, A.H., Goodin, D.S., Cree, B.A.C., Hauser, S.L., Henry, R.G., 2014. Spinal cord gray matter atrophy correlates with multiple sclerosis disability. Ann. Neurol. 76, 568–580.

Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T., 2011. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Med. Image Anal. 15, 267–282.

Shakeri, M., Tsogkas, S., Ferrante, E., Lippe, S., Kadoury, S., Paragios, N., Kokkinos, I., 2016. Sub-cortical brain structure segmentation using F-CNN's. arXiv [cs.CV].

Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2010. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. Neuroimage 49, 1524–1535.

Simard, P.Y., Steinkraus, D., Platt, J.C., Others, 2003. Best practices for convolutional neural networks applied to visual document analysis, in: ICDAR. pp. 958–962.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Sombekke, M.H., Wattjes, M.P., Balk, L.J., Nielsen, J.M., Vrenken, H., Uitdehaag, B.M.J., Polman, C.H., Barkhof, F., 2013. Spinal cord lesions in patients with clinically isolated syndrome: a powerful tool in diagnosis and prognosis. Neurology 80, 69–75.

Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B., Wang, T., 2015. Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. IEEE Trans. Biomed. Eng. 62, 2421–2433.

Souplet, J.-C., Lebrun, C., Ayache, N., Malandain, G., Others, 2008. An automatic segmentation of T2-FLAIR multiple sclerosis lesions, in: The MIDAS Journal-MS Lesion Segmentation (MICCAI 2008 Workshop). Citeseer.

Srivastava, N., Hinton, G., Krizhevsky, A., 2014. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine.

Stevenson, V.L., Leary, S.M., Losseff, N.A., Parker, G.J., Barker, G.J., Husmani, Y., Miller, D.H., Thompson, A.J., 1998. Spinal cord atrophy and disability in MS: a longitudinal study. Neurology 51, 234–238.

Stroman, P.W., Figley, C.R., Cahill, C.M., 2008. Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. Magn. Reson. Imaging 26, 809–814.

Stroman, P.W., Wheeler-Kingshott, C., Bacon, M., Schwab, J.M., Bosma, R., Brooks, J., Cadotte, D., Carlstedt, T., Ciccarelli, O., Cohen-Adad, J., Curt, A., Evangelou, N., Fehlings, M.G., Filippi, M., Kelley, B.J., Kollias, S., Mackay, A., Porro, C.A., Smith, S., Strittmatter, S.M., Summers, P., Tracey, I., 2014. The current state-of-the-art of spinal cord imaging: methods. Neuroimage 84, 1070–1081.

Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S., 2008. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. Midas J. 2008, 1–6.

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, in: Deep Learning in Medical

Image Analysis and Multimodal Learning for Clinical Decision Support. Springer International Publishing, pp. 240–248.

Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., Li, K., Mok, V.C.T., Chu, W.C.W., Wang, D., 2015. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. Biomed. Eng. Online 14, 73.

Swinnen, B., Robberecht, W., 2014. The phenotypic variability of amyotrophic lateral sclerosis. Nat. Rev. Neurol. 10, 661–670.

Thorpe, J.W., Kidd, D., Moseley, I.F., Kenndall, B.E., Thompson, A.J., MacManus, D.G., McDonald, W.I., Miller, D.H., 1996. Serial gadolinium-enhanced MRI of the brain and spinal cord in early relapsing-remitting multiple sclerosis. Neurology 46, 373–378.

Topfer, R., Foias, A., Rios, N.L., 2018. Integrated ΔB0/Rx coil array for im- proved spinal cord imaging at 3T, in: Proceedings of the 26th Annual Meeting of ISMRM. Presented at the ISMRM, p. 8930.

Topfer, R., Foias, A., Stikov, N., Cohen-Adad, J., 2018. Real-time correction of respiration-induced distortions in the human spinal cord using a 24-channel shim array. Magn. Reson. Med.

Turner, M.R., Benatar, M., 2015. Ensuring continued progress in biomarkers for amyotrophic lateral sclerosis. Muscle Nerve 51, 14–18.

Turner, M.R., Verstraete, E., 2015. What does imaging reveal about the pathology of amyotrophic lateral sclerosis? Curr. Neurol. Neurosci. Rep. 15, 45.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320.

Valsasina, P., Aboulwafa, M., Preziosa, P., Messina, R., Falini, A., Comi, G., Filippi, M., Rocca, M.A., 2018. Cervical Cord T1-weighted Hypointense Lesions at MR Imaging in Multiple Sclerosis: Relationship to Cord Atrophy and Disability. Radiology 172311.

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À., Oliver, A., Lladó, X., 2017. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159–168.

Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20, 677–688.

Vannesjo, S.J., Shi, Y., Tracey, I., Miller, K.L., Clare, S., 2017. Slice-wise first-order shimming of the human spinal cord at 7T, in: Proceedings of the 25th Annual Meeting of ISMRM. p. 5210.

Ventura, R.E., Kister, I., Chung, S., Babb, J.S., Shepherd, T.M., 2016. Cervical spinal cord atrophy in NMOSD without a history of myelitis or MRI-visible lesions. Neurol Neuroimmunol Neuroinflamm 3, e224.

Verma, T., Cohen-Adad, J., 2014. Effect of respiration on the B0 field in the human spinal cord at 3T. Magn. Reson. Med. 72, 1629–1636.

Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans. Med. Imaging 26, 405–421.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921.

Weston, J., Ratle, F., Mobahi, H., Collobert, R., 2012. Deep Learning via Semi-supervised Embedding, in: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade: Second Edition. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 639–655.

White, M.L., Zhang, Y., Healey, K., 2011. Cervical spinal cord multiple sclerosis: evaluation with 2D multi-echo recombined gradient echo MR imaging. J. Spinal Cord Med. 34, 93–98.

Xiao, Y., Fonov, V., Bériault, S., Al Subaie, F., Chakravarty, M.M., Sadikot, A.F., Pike, G.B., Collins, D.L., 2015. Multi-contrast unbiased MRI atlas of a Parkinson's disease population. Int. J. Comput. Assist. Radiol. Surg. 10, 329–341.

Yiannakas, M.C., Kearney, H., Samson, R.S., Chard, D.T., Ciccarelli, O., Miller, D.H., Wheeler-Kingshott, C.A.M., 2012. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetisation transfer measurements. Neuroimage 63, 1054–1059.

Yiannakas, M.C., Mustafa, A.M., De Leener, B., Kearney, H., Tur, C., Altmann, D.R., De Angelis, F., Plantone, D., Ciccarelli, O., Miller, D.H., Cohen-Adad, J., Wheeler-Kingshott, C.A.M.G., 2016. Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg : Application to multiple sclerosis. NeuroImage: Clinical 10, 71–77.

Yu, F., Koltun, V., 2015. Multi-Scale Context Aggregation by Dilated Convolutions. arXiv [cs.CV].

Yushkevich, P.A., Gerig, G., 2017. ITK-SNAP: An Intractive Medical Image Segmentation Tool to Meet the Need for Expert-Guided Segmentation of Complex Medical Images. IEEE Pulse 8, 54–57.

Zaimi, A., Wabartha, M., Herman, V., Antonsanti, P.-L., Perone, C.S., Cohen-Adad, J., 2018. AxonDeepSeg: automatic axon and myelin segmentation from microscopy data using convolutional neural networks. Sci. Rep. 8, 3816.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S., 2015. Conditional Random Fields as Recurrent Neural Networks, in: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1529–1537.